

Ein multiattributives Entscheidungsmodell zur Erfolgsbewertung nicht-kommerzieller Webportale

Inaugural-Dissertation
zur Erlangung des akademischen Grades eines
Doktors der Wirtschaftswissenschaft
des Fachbereichs Wirtschaftswissenschaft der
Freien Universität Berlin

vorgelegt von

Dipl.-Kfm. Daniel Delić

aus

Sinj (Kroatien)

Berlin, 31. Oktober 2008

Dekan: Univ.-Prof. Dr. Jörg Sydow

Erstgutachter: Univ.-Prof. Dr. Hans-J. Lenz

Zweitgutachterin: Univ.-Prof. Dr. Bettina Berendt

Tag der Disputation: 31.10.2008

Kurzbeschreibung

Webportale sind eine besondere Form von Websites. Sie unterscheiden sich von klassischen Websites hauptsächlich darin, dass sie Informationen aus verschiedenen anderen Internetquellen zusammenführen, aufbereiten und in einer integrierten Umgebung präsentieren. Zusätzlich beinhalten viele Portale weitere Funktionen, die sie zu Plattformen der Kommunikation und Interaktion erweitern.

Der Einsatz von Portalsoftware in nicht-kommerziellen Organisationen, wie z.B. Ämtern oder Universitäten, hat in den letzten Jahren stetig zugenommen. Diese bieten die Dienstleistungen auf ihren Portalen in der Regel kostenfrei an. Auf der anderen Seite ist der Betrieb jedoch mit Kosten verbunden. Eine möglichst optimale Gestaltung des Angebots sowie ein in der Regel beschränktes Budget sind daher wichtige Gründe für eine Evaluation des Portalerfolgs. Hierbei stellt sich das Problem, dass es im nicht-kommerziellen Bereich für die Bewertung keine Standardverfahren und -indikatoren, wie z.B. Umsatz oder Gewinn, gibt.

In der vorliegenden Arbeit wird ein Benchmarking-Ansatz in Form eines multikriteriellen bzw. multiattributiven Entscheidungsmodells vorgeschlagen, um den Erfolg eines nicht-kommerziellen Portals im Vergleich zu anderen ähnlichen Portalen zu bewerten. Multikriterielle Entscheidungsanalyse ermöglicht eine konsistente Entscheidungsfindung auf Basis der betrachteten Bewertungskriterien und Stakeholderpräferenzen. Im Fokus der Betrachtung liegen die Qualität der bereitgestellten Informationen sowie die Qualität der Informationsbereitstellung. Dazu werden das Nutzerverhalten, die Nutzermeinungen sowie die Aufwendungen zur Pflege von Inhalten analysiert und die Resultate zu einem Gesamtergebnis aggregiert.

Das Bewertungsmodell wird im Rahmen einer Fallstudie angewendet und die Ergebnisse werden diskutiert.

Abstract

Web portals are a special kind of websites. The main difference to classic websites lies in their function to gather information from various Internet sources, prepare and present it in an integrated environment. In addition, many portals offer other functions which turn them into platforms for communication and interaction.

Usage of portal software in non-commercial organizations, e.g., governmental institutions or universities, has increased steadily in the last years. Non-commercial portals offer their services for free in general. On the other hand, providing such a portal does incur costs. To offer the best service possible and usually on a limited budget are therefore important reasons for a success evaluation.

There exist no standard methods and criteria for the assessment of non-profit portal success like, e.g., turnover or profit margin. In this work a benchmarking approach in the form of a multiple-criteria respectively multi-attributive decision model is proposed for such a task. Multiple-Criteria Decision Analysis (MCDA) allows consistent decision making regarding the considered criteria and stakeholder preferences. The Focus of the analysis is the quality of provided information and the quality of information provision. Therefore user behavior, user opinions, and costs for maintaining portal content are analyzed and the results are aggregated into an overall result.

The model is applied in a case study and results are discussed.

Inhaltsverzeichnis

1 Einleitung	1
1.1 Motivation und Zielsetzung	2
1.2 Forschungsbeitrag	3
1.3 Struktur und Inhalt der Arbeit	7
2 Webportale	9
2.1 Portalkonzept	10
2.2 Klassifikation von Portalen	11
2.2.1 Portaltyp	12
2.2.2 Portalzweck	13
2.2.3 Portalausrichtung	14
2.3 Zusammenfassung	14
3 Evaluation von Informationssystemen (IS)	17
3.1 Problematik	18
3.2 Konzeption einer IS-Bewertung	19
3.2.1 Gegenstand und Aktivitätsbereich	20
3.2.2 Bewertungsperspektive	21
3.2.3 Bewertungszweck	22
3.2.4 Zeitrahmen	23
3.2.5 Analyseebenen	23
3.2.6 Bewertungsmaßstab	23
3.2.7 Methoden der Datenerfassung und -analyse	24
3.2.8 Bewertungskriterien	29
3.3 Verbreitete Ansätze zur Evaluation von Websites	43
3.3.1 Analyse der Nutzermeinungen	43
3.3.2 Analyse des Nutzerverhaltens	45
3.3.3 Expertenbasierte Analyse der Websitequalität	56
3.4 Zusammenfassung	59
4 Multikriterielle Entscheidungsanalyse (MCDA)	61
4.1 Grundmodell	62
4.2 Verfahren	66

4.2.1	Verfahren mit Präferenzfunktional	67
4.2.2	Verfahren ohne Präferenzfunktional	68
4.3	Nutzwertanalyse	68
4.4	Zusammenfassung	70
5	Bewertungsmodell für Webportale	73
5.1	Bewertungsstrategie	74
5.2	Evaluationsansatz	76
5.2.1	Gegenstand und Aktivitätsbereich	76
5.2.2	Bewertungsperspektive	77
5.2.3	Bewertungszweck	78
5.2.4	Zeitrahmen	78
5.2.5	Bewertungsmaßstab	78
5.2.6	Analyseebenen	79
5.2.7	Methoden der Datenerfassung und -analyse	79
5.2.8	Bewertungskriterien	79
5.3	Entscheidungsmodell	95
5.3.1	Nutzerverhalten	95
5.3.2	Nutzermeinungen	96
5.3.3	Aufwendungen	98
5.3.4	Gesamtbewertung	98
5.4	Zusammenfassung	99
6	Fallstudie Bildungsserver	101
6.1	Betrachtete Portale	102
6.2	Bewertung des Nutzerverhaltens	103
6.2.1	Datenaufbereitung	104
6.2.2	Ergebnisse und Interpretation	108
6.2.3	Diskussion	115
6.3	Bewertung der Nutzermeinungen	116
6.3.1	Datenaufbereitung	117
6.3.2	Ergebnisse und Interpretation	124
6.3.3	Diskussion	131
6.4	Bewertung Aufwendungen	132
6.4.1	Datenaufbereitung	132
6.4.2	Ergebnisse und Interpretation	133
6.4.3	Diskussion	134
6.5	Gesamtbewertung	135
6.5.1	Ergebnisse und Interpretation	135
6.5.2	Diskussion	137
6.6	Schlussfolgerungen für den Portalbetrieb	138
6.7	Zusammenfassung	139

7 Zusammenfassung und Ausblick	141
7.1 Überblick	142
7.2 Schlussfolgerungen	143
7.2.1 Nutzerverhalten	143
7.2.2 Nutzermeinungen	144
7.2.3 Aufwendungen	144
7.2.4 Gesamtbewertung	145
7.3 Ausblick	145
Literaturverzeichnis	147
A Ergänzungsansätze bei unvollständigen Daten	171
A.1 Fehlmechanismen in Daten	171
A.2 Ergänzungsmethoden	172
A.3 Imputation mit SAS (PROC MI)	175
B Histogramme zu den Umfrageergebnissen	177
C Validierungsergebnisse zum Pfadmodell	181
C.1 Signifikanz der Faktorladungen	182
C.2 Faktorladungen (<i>Crossloadings</i>)	187
C.3 AVE- und Korrelationsmatrizen	192
C.4 Abhängigkeiten zwischen den Konstrukten	197
C.5 Bestimmtheitsmaß R^2	202
C.6 Cronbachs α	203

Abbildungsverzeichnis

2.1	Schema des technischen Aufbaus eines Webportals	11
2.2	Konzepthierarchien zur Portalklassifikation	12
3.1	IS-Evaluationspyramide	21
3.2	Stakeholdertypologie	22
3.3	Aus den betrachteten Ansätzen synthetisierte Klassifikation von Methoden zur Datenerfassung und -analyse	24
3.4	Aus den betrachteten Ansätzen synthetisierte Klassifikation von IS-Bewertungskriterien	30
3.5	<i>Technology Acceptance Model</i>	40
3.6	<i>IS Success Model</i>	41
3.7	<i>Task-Technology Fit</i> -Modell	42
3.8	Ansätze zur Analyse von Nutzermeinungen	44
3.9	Taxonomie des Web Mining	47
3.10	Web Usage Mining Prozess	49
3.11	Anwendungsgebiete des Web Usage Mining	53
3.12	Aus den betrachteten Ansätzen synthetisierte Klassifikation von Bewertungskriterien zur Webusability-Inspektion	58
4.1	Beispiel einer dreistufigen Zielhierarchie	63
4.2	Beispielschema einer Ergebnismatrix E	64
4.3	Beispielschema einer Entscheidungsmatrix D	66
4.4	Verfahren der Mehrzielanalyse	66
5.1	Konzepthierarchien zur Analyse des Nutzerverhaltens	81
5.2	Effektivitäts-Effizienz-Matrix	86
5.3	Vereinfachtes D/M92	87
5.4	Pfadmodell zur Erfassung der Nutzermeinungen	89
5.5	Zielhierarchie für Webportalerfolg	96
6.1	Homepage des Schweizerischen Bildungsservers vom 14.05.2008 . .	103
6.2	Output des J48-Algorithmus für Portal P2	106
6.3	MINT-Abfrage zur Extraktion der von der Homepage (Typ H) zu Seiten des Typs M verlaufenden Klickpfade	108

6.4	Präferenz bezüglich Effektivität und Effizienz von Zugriffen	109
6.5	Anteile an Sessions mit aufeinander folgenden NINE-Abrufen	114
6.6	Histogramme zu den erhobenen Daten für die Items bzw. Variablen INFQ3 und VERT2 der Portale P1–P4	121
6.7	Respezifiziertes Pfadmodell zu den Nutzermeinungen	124
6.8	Prozentuale Verteilung der Nutzergruppen	125
6.9	Prozentuale Verteilung der Nutzungshäufigkeiten	126
6.10	Prozentuale Verteilung der angegebenen Nutzungsgründe	126
6.11	Clusterergebnis auf Basis der demographischen Daten	127
A.1	Aufruf der Prozedur PROC MI zur Monotonisierung der Muster der fehlenden Werte in der Datentabelle mittels MCMC-Simulation . .	176
A.2	Aufruf der Prozedur PROC MI zur Imputation fehlender Werte in der Datentabelle mittels logistischer Regression	176
B.1	Histogramme zu den erhobenen Daten der Portale P1–P4 (1)	177
B.2	Histogramme zu den erhobenen Daten der Portale P1–P4 (2)	178
B.3	Histogramme zu den erhobenen Daten der Portale P1–P4 (3)	179
B.4	Histogramme zu den erhobenen Daten der Portale P1–P4 (4)	180

Tabellenverzeichnis

3.1	Modelle relevanter Variablen von IS-Erfolg	39
3.2	Verbreitete Ansätze zur Evaluation von Websites	43
3.3	Ansätze die auf dem D/M92 bzw. D/M03 basieren	45
3.4	Ansätze die von DeLone u. McLean identifizierte IS-Erfolgskategorien bzw. -kriterien verwenden	46
3.5	Web-Begriffe	48
3.6	Wichtige Felder eines Webserver-Logs	50
3.7	Data Mining Methoden	52
3.8	Nielsens 10 Usability-Design-Prinzipien	57
3.9	Ansätze zur Usability-Inspektion von Websites	59
5.1	Betrachtete Servicetypen	82
5.2	Betrachtete Seitentypen	83
5.3	Betrachtete Referrertypen	83
5.4	Hypothesen zu Konstruktabhängigkeiten	90
5.5	Fragebogen	92
5.6	Demographischer Teil des Fragebogens	93
5.7	Beispieldausschnitt der quartilbasierten Ergebnismatrix zur Komponente „Nutzermeinungen“	97
5.8	Beispieldausschnitt der quartilbasierten Entscheidungsmatrix zur Komponente „Nutzermeinungen“	97
6.1	H-Sessions	104
6.2	Hyperlink-zu-Text-Verhältnisse der verschiedenen Seitentypen	105
6.3	Klassifikationsgütekennzahlen	107
6.4	Präferenzen zu den Seitentypen und zum Typ DOC	109
6.5	Häufigkeiten und Anteile von Service-Anfragen	109
6.6	Häufigkeiten von Zugriffen auf Content-Seiten	110
6.7	Häufigkeiten von Zugriffen auf Navigations-Seiten	110
6.8	Durchschnittliche Pfadlängen in den H-Sessions	110
6.9	Entscheidungsmatrix zum Nutzerverhalten	111
6.10	Häufig vorkommende Pfadmuster für Zugriffe auf MNINE-Seiten	113
6.11	Statistik zur Umfrageteilnahme	117

6.12	Anteile fehlender Werte	118
6.13	Beispiel eines monotonen Musters fehlender Werte	119
6.14	Indikatoren für Konstruktvalidität	122
6.15	Indikatoren für strukturelle Validität	122
6.16	Anteile der Befragten Nutzer mit mehr als 12 Monaten Portalerfahrung und mehr als fünf Jahren Weberfahrung	127
6.17	Präferenzen zu den individuellen Auswirkungen	128
6.18	Quartile zum Umfrageergebnis für P2	129
6.19	Entscheidungsmatrix zu den Nutzermeinungen	130
6.20	Contentänderungsverhältnisse pro Session (Werte in Klammern) auf Basis der ermittelten wöchentlichen Anzahl an Sessions und der geschätzten wöchentlichen Contentänderungsvorgänge durch den Portalbetreiber	133
6.21	Präferenzen zu den Aufwendungen	133
6.22	Entscheidungsmatrix zu den Aufwendungen	134
6.23	Präferenzen zur Gesamtbewertung	136
6.24	Entscheidungsmatrix zur Gesamtbewertung	136
C.1	t-Statistik zu den gepoolten Faktorladungen (\bar{Q}) für P1	183
C.2	t-Statistik zu den gepoolten Faktorladungen (\bar{Q}) für P2	184
C.3	t-Statistik zu den gepoolten Faktorladungen (\bar{Q}) für P3	185
C.4	t-Statistik zu den gepoolten Faktorladungen (\bar{Q}) für P4	186
C.5	Gepoolte Faktorladungsmatrix zu P1	188
C.6	Standardabweichungen ($\sqrt{\hat{B}}$) zu Tabelle C.5 für P1	188
C.7	Gepoolte Faktorladungsmatrix zu P2	189
C.8	Standardabweichungen ($\sqrt{\hat{B}}$) zu Tabelle C.7 für P2	189
C.9	Gepoolte Faktorladungsmatrix zu P3	190
C.10	Standardabweichungen ($\sqrt{\hat{B}}$) zu Tabelle C.9 für P3	190
C.11	Gepoolte Faktorladungsmatrix zu P4	191
C.12	Standardabweichungen ($\sqrt{\hat{B}}$) zu Tabelle C.11 für P4	191
C.13	Gepoolte AVE-Werte (\bar{Q}) für P1	193
C.14	Gepoolte AVE- und Korrelationsmatrix für P1	193
C.15	Standardabweichungen ($\sqrt{\hat{B}}$) zu Tabelle C.14 für P1	193
C.16	Gepoolte AVE-Werte (\bar{Q}) für P2	194
C.17	Gepoolte AVE- und Korrelationsmatrix für P2	194
C.18	Standardabweichungen ($\sqrt{\hat{B}}$) zu Tabelle C.17 für P2	194
C.19	Gepoolte AVE-Werte (\bar{Q}) für P3	195
C.20	Gepoolte AVE- und Korrelationsmatrix für P3	195
C.21	Standardabweichungen ($\sqrt{\hat{B}}$) zu Tabelle C.20 für P3	195
C.22	Gepoolte AVE-Werte (\bar{Q}) für P4	196
C.23	Gepoolte AVE- und Korrelationsmatrix für P4	196
C.24	Standardabweichungen ($\sqrt{\hat{B}}$) zu Tabelle C.23 für P4	196

C.25 t-Statistik zu den gepoolten Pfadkoeffizienten (\bar{Q}) für P1	198
C.26 t-Statistik zu den gepoolten Pfadkoeffizienten (\bar{Q}) für P2	199
C.27 t-Statistik zu den gepoolten Pfadkoeffizienten (\bar{Q}) für P3	200
C.28 t-Statistik zu den gepoolten Pfadkoeffizienten (\bar{Q}) für P4	201
C.29 Gepoolte R^2 -Werte (\bar{Q}) für P1	202
C.30 Gepoolte R^2 -Werte (\bar{Q}) für P2	202
C.31 Gepoolte R^2 -Werte (\bar{Q}) für P3	202
C.32 Gepoolte R^2 -Werte (\bar{Q}) für P4	202
C.33 Gepoolte α -Werte (\bar{Q}) für P1	203
C.34 Gepoolte α -Werte (\bar{Q}) für P2	203
C.35 Gepoolte α -Werte (\bar{Q}) für P3	204
C.36 Gepoolte α -Werte (\bar{Q}) für P4	204



Abkürzungsverzeichnis

AHP	<i>Analytic Hierarchy Process</i>
AVE	<i>Average Variance Extracted</i>
BI	<i>Business Intelligence</i>
DEA	<i>Data Envelopment Analysis</i>
D/M92	<i>IS Success Model</i>
D/M03	<i>IS Success Model 2003</i>
ELECTRE	<i>ELiminacion Et Choix Traduisant la REqualite</i>
FP	<i>False Positive</i>
HTML	<i>Hypertext Markup Language</i>
HTTP	<i>Hypertext Transfer Protocol</i>
IMGP	<i>Interactive Multiple Goal Programming</i>
IP	<i>Internet Protocol</i>
IQ	Informationsqualität
IS	Informationssystem
LINMAP	<i>LINear Programming techniques for Multidimensional Analysis of Preference</i>
MAR	<i>Missing at random</i>
MAUT	<i>Multi-Attribute Utility Theory</i>
MCAR	<i>Missing completely at random</i>
MCDA	<i>Multiple-Criteria Decision Analysis</i>
MCMC	<i>Markov Chain Monte Carlo</i>
MDS	Multidimensionale Skalierung
NMAR	<i>Not missing at random</i>
NWA	Nutzwertanalyse

PEOU	<i>Perceived Ease of Use</i>
PLS	<i>Partial Least Squares</i>
PROMETHEE	<i>Preference Ranking Organization METHods for Enrichment Evaluations</i>
PU	<i>Perceived Usefulness</i>
TAM	<i>Technology Acceptance Model</i>
TCO	<i>Total Cost of Ownership</i>
TP	<i>True Positive</i>
TPC	<i>Technology-to-Performance Chain</i>
TRA	<i>Theory of Reasoned Action</i>
TTF	<i>Task-Technology Fit</i>
URI	<i>Uniform Resource Identifier</i>
WAP	<i>Wireless Application Protocol</i>
WBIS	<i>Web-basiertes Informationssystem</i>
WUM	<i>Web Utilization Miner</i>
WWW	<i>World Wide Web</i>

Symbolverzeichnis

\bar{Q}	Mittelwert der m Punktschätzer für Q
\bar{U}	Mittelwert der m Varianzschätzer für \bar{Q}
$\hat{\lambda}$	Schätzer für den Anteil der fehlenden Information über Q
\hat{B}	Schätzer für die Varianz zwischen den m Punktschätzern für Q
\hat{Q}	Punktschätzer für Q
\hat{T}	Schätzer für die totale Varianz von \bar{Q} über m Imputationen
\hat{U}	Varianzschätzer von \hat{Q}
\mathcal{P}	Menge der betrachteten Zielseitentypen
\mathcal{R}	Menge der betrachteten Referrertypen
\mathcal{S}	Menge der betrachteten Besucher- bzw. Nutzersessions
\mathcal{S}^{ACT}	Menge der aktiven Sessions
\mathcal{S}^{POS}	Menge der positiven Sessions
\mathcal{Z}	Menge der betrachteten Ziel- bzw. Webserverdokumente
$\mathfrak{a} \sqsubset \mathfrak{b}$	\mathfrak{a} ist eine Subsequenz von \mathfrak{b}
$\mathfrak{v}(\cdot)$	Wertfunktion, die eine Ordnungsrelation in eine numerische Beziehung überträgt
m_{sz}	Anzahl der Elemente in einer z -Subsequenz in Session s
r	Zugriff (Request) auf ein Ziel- bzw. Webserverdokument in einer Session
s	Menge der Requests in einer Session
z	Ziel- bzw. Webserverdokumententyp aus der Menge $\mathcal{Z} = \{\mathcal{P}, \text{DOC}\}$
$z\text{-}seq(s)$	z -Subsequenz von $seq(s)$ in Session s

ν	für eine t-Statistik aus m Imputationen errechnete Anzahl an Freiheitsgraden
$\Phi(a)$	Mehrzielpräferenzfunktion zur Bestimmung des Rankings der Alternative a auf Basis der Ausprägungen der betrachteten Kriterien
A	Menge aller betrachteten Alternativen
a	eine Alternative aus der Menge A der betrachteten Alternativen
$a' \sim a''$	zwischen a' und a'' liegt Indifferenz vor
$a' \succsim a''$	a' wird gegenüber a'' schwach präferiert
C	Menge der betrachteten Zielkriterien bzw. Zielattribute
c	ein Zielkriterium aus der Menge C der betrachteten Zielkriterien
$cav(vtyp)$	Contentänderungsverhältnis für den Vorgangstyp $vtyp$
$con(r)$	dem Request r zugewiesenes Konzept bzw. der Typ des Requests
$eff(z)$	Effektivität von z -Requests
$efz(z)$	durchschnittliche Effizienz von z -Requests
$length(z-sseq(s))$	Länge einer z -Subsequenz von $seq(s)$
m	Anzahl der Imputationen
Q	Statistik der vollständigen Daten (Mittelwert, Varianz usw.)
r	relativer Anstieg der Varianz einer Variable durch die Imputation
R^2	Bestimmtheitsmaß; Anteil der Varianz einer endogenen latenten Variable, welcher durch exogene latente Variablen bestimmt wird
RE	relative Effizienz einer finiten Anzahl von m Imputationen im Vergleich zu $m \rightarrow \infty$
$refcon(r)$	dem Refferer zu Request r zugewiesenes Konzept bzw. der Typ des Referrers
$seq(s)$	Sequenz der Requests in Session s
$sseq(s)$	Subsequenz von $seq(s)$ in Session s

$u(x_j(a))$	ordinale Nutzenfunktion, die die Ausprägung x eines Zielkriteriums c_j der Alternative a in einen Nutzenwert u überführt
$U^{\text{NWA}}(a)$	mittels Nutzwertanalyse ermittelter Präferenzindex bzw. <i>utility score</i> für Alternative a
V	Menge aller Vorgänge zur Änderung von Portalcontent
v	ein Vorgang zur Änderung von Portalcontent (<i>create</i> , <i>update</i> oder <i>delete</i>)
V_{vtyp}	Menge aller Vorgänge des Typs $vtyp$ zur Änderung von Portalcontent
$VTyp$	Menge aller Vorgangstypen zur Änderung von Portalcontent
$vtyp$	ein Vorgangstyp zur Änderung von Portalcontent
w	Gewichtungsfaktor für ein Zielkriterium c
X	Ausprägungsbereich (Domäne) eines Zielkriteriums
x	Ausprägung eines Zielkriteriums aus dem Ausprägungsbereich X
AVE	Anteil der durchschnittlichen Varianz in den Messvariablen, welcher von deren zugehörigem Konstrukt erfasst wird, im Verhältnis zur Varianz der Messfehler
Cronbachs α	durchschnittliche Inter-Korrelation unter den Messvariablen eines Konstrukts (bzw. einer latenten Variable)
DBL	Servicetyp DBL (Suchergebnisliste einer Datenbank-Abfrage)
DOC	Servicetyp DOC (herunterladbares Dokument)
H	Webseitentyp H (Homepage bzw. betrachtete Startseite)
INFQ1–INFQ5	Fragen zur Fragebogenkategorie ‘Informationsqualität’. Diese werden auch als <i>Items</i> bezeichnet. Zusammen bilden diese die Kategorie INFQ. Im Rahmen der Imputation stellen diese die betrachteten Variablen dar. Im Rahmen der Strukturgleichungsmodellierung stellen diese die Messvariablen dar und die Kategorie INFQ stellt dort das Konstrukt bzw. die latente Variable dar. Im Rahmen des Entscheidungsmodells stellen diese die (Ziel-) Attribute bzw. (Ziel-) Kriterien dar.
L	Servicetyp L (Suchergebnisliste einer Suchmaschine)

LOYA1, LOYA2	Fragen zur Fragebogenkategorie ‘Loyalität’. Siehe hierzu INFQ1–INFQ5.
M	Webseitentyp M (reine Content-Seite)
MNE	Webseitentyp MNE (Content-Navigations-Seite portalextern)
MNI	Webseitentyp MNI (Content-Navigations-Seite portalintern)
MNINE	Webseitentyp MNINE (Content-Navigations-Seite portalintern und -extern)
NE	Webseitentyp NE (Navigations-Seite portalextern)
NI	Webseitentyp NI (Navigations-Seite portalintern)
NINE	Webseitentyp NINE (Navigations-Seite portalintern und -extern)
RISK1, RISK2	Fragen zur Fragebogenkategorie ‘Risiko’. Siehe hierzu INFQ1–INFQ5.
SYSQ1–SYSQ6	Fragen zur Fragebogenkategorie ‘Systemqualität’. Siehe hierzu INFQ1–INFQ5.
VERT1, VERT2	Fragen zur Fragebogenkategorie ‘Vertrauen’. Siehe hierzu INFQ1–INFQ5.
WISS1, WISS2	Fragen zur Fragebogenkategorie ‘Wissen’. Siehe hierzu INFQ1–INFQ5.
ZUFR1, ZUFR2	Fragen zur Fragebogenkategorie ‘Zufriedenheit’. Siehe hierzu INFQ1–INFQ5.

Kapitel 1

Einleitung

Seit der Entstehung des *World Wide Web* (WWW) in den 80-ern ist die Anzahl an Websites bis Juni 2008 auf geschätzte 172 Mio. (Netcraft.com 2008) angestiegen. Eine ursprünglich aus Suchdiensten wie Yahoo (www.yahoo.com) weiterentwickelte Form von Website ist das Webportal.

Webportale unterscheiden sich von klassischen Websites hauptsächlich darin, dass sie Informationen aus verschiedenen anderen Internetquellen zusammenführen, aufbereiten und in einer integrierten Umgebung präsentieren. Zusätzlich beinhalten viele Portale Online-Tools, durch welche sie zu Plattformen der Produktivität (z. B. eMail, Kalender oder Adressbuch), Kommunikation (z. B. Foren oder Chat) und Zusammenarbeit (z. B. Groupware) erweitert werden (Butters 2003).

Wegen ihrer speziellen Charakteristik haben Webportale eine zunehmende Bedeutung im WWW. Gartner schätzt das Marktvolumen 2006 für Portale, Prozesssoftware und Middleware allein im Raum Europa, Mittlerer Osten und Afrika auf rund \$3,8 Mrd., was einem Wachstum von 16% im Vergleich zu 2005 entspräche (Gartner 2007).

Neben dem zunehmenden Einsatz im kommerziellen Bereich nutzen auch nicht-kommerzielle Organisationen verstärkt Portale für die Ausgestaltung ihrer Webpräsenzen. Beispiele hierfür sind institutionelle Einrichtungen wie Behörden (z. B. www.finanzamt.de) oder Hochschulen (z. B. www.fu-berlin.de) sowie private Organisationen (z. B. www.linux.org). Die Portale werden u. a. mit dem Ziel eingesetzt, den Nutzern informative Inhalte und hilfreiche Funktionen bereit zu stellen. So können sie beispielsweise von Besuchern dazu genutzt werden, Informationen einzuholen oder Transaktionen online durchzuführen (z. B. Abgabe der Steuererklärung).

Der Gebrauch nicht-kommerzieller Portale ist, wie der Name bereits andeutet, für den Nutzer in der Regel kostenfrei, d. h. Ziel des Portalbetriebs ist nicht die Erwirtschaftung von Gewinnen. Auf der anderen Seite kann der Betrieb je nach Größe eines solchen Portals und dem damit verbundenen Administrationsaufwand erhebliche Kosten verursachen. Diese Ausgaben zu rechtfertigen und Möglichkeiten aufzudecken, das Angebot eventuell zu verbessern, sind zusammen mit einem in

der Regel beschränkten Budget wichtige Gründe aus denen sich für Betreiber nicht-kommerzieller Portale der Bedarf an einer Erfolgsbewertung ergibt.

Das Prinzip von Angebotserstellung und Nachfrage kann analog zu kommerziellen Portalen betrachtet werden. Ein Produkt bzw. eine Dienstleistung wird dem Nutzer („Konsumenten“) angeboten und dieser wird es in Abhängigkeit der Qualität, des Nutzens und des zu zahlenden Preises mehr oder weniger stark nachfragen. Eine starke Produktnachfrage ist in diesem Zusammenhang als Erfolg zu werten, wenn der Aufwand den „Ertrag“ nicht übersteigt.

Ein Produktpreis kann im Gegensatz zu kommerziellen Portalen allerdings nicht in Form eines Geldbetrags erfasst werden. Finanzielle Indikatoren wie Umsatz oder Gewinn, die im kommerziellen Bereich gebräuchlich sind und eindeutige Erfolgsaussagen ermöglichen, lassen sich somit nicht verwenden.

In dem hier beschriebenen Rahmen werden daher für den nicht-kommerziellen Kontext geeignete Kriterien benötigt, die eine sinnvolle Bewertung erlauben. Die Entwicklung eines entsprechenden Ansatzes ist Gegenstand dieser Arbeit.

1.1 Motivation und Zielsetzung

Ein Webportal ist ein bestimmter Typ von Informationssystem (IS) – im Folgenden werden sowohl der Begriff „Informationssystem“ als auch der Begriff „Informationsysteme“ mit „IS“ abgekürzt. Ein IS ist ein computerbasiertes Mensch-Maschine-System zur Bereitstellung und Verarbeitung von Informationen (Davis u. Olson 1985, S. 6). Neben allgemeinen Basis-Charakteristiken kann ein IS je nach Einsatzgebiet und -zweck weitere spezielle Eigenschaften und Funktionen besitzen. Da die Interaktion mit den Nutzern Teil dieses „Systems“ ist, beeinflusst auch die humane „Komponente“ dessen Eigenschaften, Funktionsweisen und Auswirkungen. Aufgrund dieser Konstellation ergibt sich ein großes Feld an möglichen Ansatzpunkten zur Bewertung. Sowohl qualitative Methoden aus der Sozialforschung, wie z. B. die Beobachtung von Phänomenen, als auch rein quantitative, an die Verfahrensweise in den Naturwissenschaften angelehnte Methoden, wie z. B. das Laborexperiment, haben hier ihre Berechtigung.

Einige Forscher argumentieren, dass die Ursache für dieses große Feld an Ansatzmöglichkeiten in einer bis heute unklaren Sichtweise auf die „Abhängige Variable“ (Keen 1980), d. h. dem ultimativen Erfolgsbewertungskriterium für IS, liegt (Benbasat u. Zmud 2003; Cushing 1990; Keen 1980). Vermutlich aus dieser Unklarheit heraus existiert eine Vielzahl an Bewertungsansätzen, in denen unterschiedlichste Bereiche eines IS fokussiert und mittels verschiedenster Methoden analysiert werden. Es gibt daher keine allgemein gültigen Standardvorgehensweisen.

Aufgrund dieser Unklarheit sowie der Vielfalt vorhandener Ansätze, wird hier die Ansicht vertreten, dass der derzeit beste Weg für die Entwicklung eines Evaluationsmodells für nicht-kommerzielle Webportale in einer adäquaten Auswahl der zu untersuchenden Komponenten und (Komponenten-)Bereiche eines Systems und

damit einhergehend in einer entsprechenden Zusammenstellung adäquater Bewertungsmethoden und -kriterien besteht. Hierzu sind vier grundlegende Punkte zu klären:

1. **Systemabgrenzung:** Voraussetzung für die Entwicklung eines Bewertungsansatzes ist eine klare Abgrenzung des zu analysierenden Systems. Es muss eindeutig definiert werden, aus welchen Komponenten das System besteht bzw. welche Komponenten für eine Bewertung relevant sind. Wie oben erläutert, ist ein IS ein System aus dem Zusammenspiel von Mensch und Maschine. Während die technischen Komponenten relativ leicht zu identifizieren sind, ist es schwieriger, die für eine Bewertung relevanten Stakeholdergruppen zu bestimmen und einzubeziehen. Mindestens eine relevante Gruppe stellen die Nutzer dar.
2. **Betrachtete Systembereiche:** Liegt ein klar abgegrenztes System vor, ist festzulegen, welche Bereiche innerhalb der vorliegenden Systemkomponenten relevant sind und im Rahmen der Bewertung betrachtet werden. Das kann z. B. die Qualität der produzierten Informationen durch das System oder die Zufriedenheit der Nutzer sein.
3. **Datenerfassung und -analyse:** Auf Basis der zu betrachtenden Bereiche ist festzulegen, welche Methoden zur Datenerfassung sowie welche Bewertungskriterien jeweils eingesetzt werden sollen.
4. **Bewertung:** Um letztlich vernünftige Schlüsse ziehen zu können, ist zu klären, wie die Auswertungsergebnisse zu interpretieren sind. Es müssen dazu ein Bewertungsverfahren sowie ein Bewertungsmaßstab festgelegt werden.

Zielsetzung dieser Arbeit ist die Entwicklung eines Modells zur Bewertung nicht-kommerzieller Webportale, unter Berücksichtigung der obigen vier grundlegenden Punkte. Im Fokus liegen dabei die Betrachtung der Qualität der bereitgestellten Informationen sowie die Qualität der Informationsbereitstellung. Das Modell ist somit insbesondere für nicht-kommerzielle *Informationsportale* konzipiert. Die einzusetzenden Verfahren zur Datenerfassung und -analyse sollen außerdem ohne großen Aufwand anwendbar sein, da Betreibern nicht-kommerzieller Portale in der Regel für entsprechende Untersuchungen nur ein beschränktes Budget zur Verfügung steht.

1.2 Forschungsbeitrag

Der Forschungsbeitrag dieser Arbeit liegt in der Entwicklung eines Bewertungsmodells für nicht-kommerzielle Portale, welches verschiedene bestehende Verfahren aus der IS-Forschung integriert und so eine umfassende, multiperspektivische Be trachtungsweise des zu evaluierenden Systems ermöglicht (siehe Delić 2007). Dieser

Ansatz erlaubt die bessere Beurteilung eines Portals, als es mittels eines einzelnen Verfahrens möglich wäre.

Wie bereits oben angeführt, gibt es bisher keine Standardmethoden und -kriterien zur Erfolgsbewertung von IS. Die hier im vorgestellten Modell eingesetzten Verfahren basieren auf häufig in der Literatur verwendeten und akzeptierten Ansätzen.

Je nach Relevanz ist ein einzelnes Kriterium in der Gesamtbetrachtung mehr oder weniger ausschlaggebend für die Erfolgsbewertung. Durch das Fehlen von Standards, hängt die Einstufung der Relevanz von den Präferenzen des jeweiligen Stakeholders ab. Verschiedene Stakeholder können je nach Sichtweise zu verschiedenen Schlüssen bezüglich des Erfolgs eines Portals kommen. Das hier vorgeschlagene Modell berücksichtigt diesen Umstand, indem individuelle Präferenzen in die Bewertung einbezogen werden.

Des Weiteren gibt es auch keine Standards bezüglich der Bewertung bzw. der Interpretation der Ergebnisse. Der hier im Rahmen des entwickelten Modells umgesetzte Vorschlag sieht vor, durch den Vergleich mehrerer in ihren Eigenschaften ähnlicher Portale eine Basis zur Interpretation von Ergebnissen zu schaffen. Je mehr Portale in die Analyse einbezogen werden, desto stärker kann diese Basis als „Standard“ gewertet werden.

Im Folgenden wird der in dieser Arbeit entwickelte Ansatz anhand der im vorangegangenen Abschnitt zur Problemstellung aufgeführten vier Punkte detaillierter dargestellt und Unterschiede zu bestehenden Ansätzen werden hervorgehoben:

1. **Systemabgrenzung:** In den meisten bekannten Ansätzen zur Evaluation von Websites besteht das betrachtete System in der Regel aus der Website als technische und den Nutzern als humane Komponente (siehe Abschnitt 3.3). Effekte, die aus dem Zusammenspiel von Website und Nutzer entstehen, sind dort Gegenstand der Analyse – unter dem Begriff „Effekte“ werden hier sowohl vom Nutzer wahrgenommene Merkmale einer Site, z. B. die Informationsqualität, als auch Auswirkungen der Nutzung, z. B. erweitertes Wissen, verstanden.

Man kann davon ausgehen, dass Bereitstellungsprozesse einen wesentlichen Einfluss auf die Qualität einer Website haben. So hängt z. B. die Aktualität von bereitgestellten Informationen unter anderem von der Aktualisierungsfrequenz der Inhalte durch den Betreiber ab. Um ein möglichst vollständiges Bild des Erfolgs eines Portals zu erhalten, ist es daher sinnvoll, in Zusammenhang mit Effekten auch Bereitstellungsprozesse zu betrachten. In dem hier betrachteten Rahmen wird daher neben den Nutzern auch der Betreiber als eine weitere wichtige Stakeholdergruppe aufgenommen.

2. **Betrachtete Systembereiche:** Viele Evaluationsansätze betrachten die aus dem Zusammenspiel von System und Nutzer resultierenden Effekte anhand der Analyse des Nutzerverhaltens oder der Nutzermeinungen. Ein weiterer verbreiteter Ansatz ist die Bewertung von Websitequalität durch Experten

anhand „objektiver“ Kriterien. In diesem Fall nimmt der Experte die Rolle eines Nutzers ein (siehe Abschnitt 3.3).

Da Nutzerverhalten und Nutzermeinungen als Erfolgsindikatoren auf breite Akzeptanz in der IS-Forschung stoßen (siehe Abschnitt 3.2.8), beinhaltet das entwickelte Bewertungsmodell unter anderem diese zwei Bereiche (siehe Delić u. Lenz 2008a, b). Die expertenbasierte Analyse wird nicht miteinbezogen, da diese Nutzerverhalten und Nutzermeinungen nur simuliert, ohne die wahren Reaktion zu kennen. Diese können, müssen aber nicht, mit dem Urteil des Experten übereinstimmen.

Ein weiterer Bereich, der sich aus der Systemabgrenzung ergibt und hier betrachtet wird, sind Bereitstellungsprozesse auf Seiten des Betreibers. Ziel ist es, wie oben beschrieben, eine erweiterte Betrachtungsperspektive zu erhalten.

3. **Datenerfassung und -analyse:** Die Auswahl der Verfahren zur Datenerfassung und -analyse orientiert sich an gängigen und allgemein akzeptierten Ansätzen (siehe Abschnitt 3.2.8), unter Berücksichtigung des hier festgelegten Grundsatzes, dass die Umsetzung mit relativ geringem (u. a. finanziellen) Aufwand realisierbar sein soll.

Zur Erfassung des Nutzerverhaltens werden die von einem Portal erzeugten Logdaten ausgewertet. Die Auswahl der Kriterien zielt darauf ab, das Interesse an und die Erreichbarkeit von Informationsinhalten zu analysieren.

Zur Erfassung der Nutzermeinungen wird ein strukturierter Fragebogen eingesetzt. Die Auswahl der Kriterien bzw. Fragen zielt primär auf eine Analyse der von den Nutzern individuell wahrgenommenen Effekte ab.

Die Erfassung von Bereitstellungsprozessen erfolgt mittels Befragung des Portaltreibers. Gegenstand der Kriterienauswahl ist die Analyse von Aufwendungen zur Pflege der bereitgestellten Informationsinhalte.

4. **Bewertung:** Als Bewertungsverfahren wird die komparative Analyse bzw. Benchmarking eingesetzt, d. h. der Vergleich der Ergebnisse mehrerer Portale bildet die Interpretationsgrundlage bzw. den Maßstab zur Bewertung. Die Bewertungskriterien entsprechen hierbei den zur Datenerfassung verwendeten. Voraussetzung für einen Vergleich ist, dass diese quantitativen Typs sind.

Wie oben erwähnt, gibt es keine Standards zur Erfolgsbewertung nicht-kommerzieller Portale. Die verwendeten Kriterien, die Einschätzung der relativen Wichtigkeit eines Kriteriums sowie die Interpretation der Ergebnisse können je nach den Präferenzen einer Stakeholdergruppe variieren. Um den Bewertungsprozess an die Bedürfnisse der jeweiligen Stakeholdergruppe besser anpassen zu können, werden entsprechende Präferenzen in das Bewertungsmodell mit aufgenommen. Im Rahmen des hier durchgeföhrten Bewertungsansatzes wird dazu eine spezielle Form der komparativen Analyse – die Nutz-

wertanalyse – eingesetzt. Diese gehört zu den Methoden der Multikriteriellen Entscheidungsanalyse (engl.: *Multiple-Criteria Decision Analysis* (MCDA)), welche eine konsistente Entscheidungsfindung unter den gegebenen Kriterien und Präferenzen ermöglichen.

Die Idee, den Erfolg einer Website bzw. eines Portals mittels komparativer Analyse zu erfassen, ist nicht neu. In der Literatur finden sich verschiedene Arbeiten, die auf Basis qualitativer und/oder quantitativer Bewertungskriterien eine Evaluation durchführen. Im Folgenden wird ein kurzer Überblick zu bestehenden Ansätzen gegeben, Unterschiede zu dem in dieser Arbeit entwickelten Modell werden aufgezeigt.

- **Qualitative Bewertungskriterien:** Eine qualitative Analyse ist sehr hilfreich bei der Erfassung reichhaltiger, multiperspektivischer Informationen zum Evaluationsgegenstand und dessen Kontext (Abschnitt 3.2.7). Ein komparativer Vergleich ist jedoch durch die Nicht-Quantifizierbarkeit bzw. nicht vorgenommene Quantifizierung der Analyseergebnisse nur schwer möglich. Entsprechende Ansätze (z.B. Oates 2002) eignen sich daher nur bedingt für Benchmarking bzw. Multikriterielle Entscheidungsanalyse.
- **Expertenbasierte Analyse von Websitequalität:** Bei der Bewertung durch Experten wird das Vorhandensein objektiv erfassbarer Eigenschaften überprüft und/oder die Qualität bestehender Eigenschaften eingestuft. Die expertenbasierte Analyse kann manuell (z.B. Kaylor u. a. 2001; Misic u. Johnson 1999) oder automatisiert (z.B. Jutla u. a. 1999; Olsina u. Rossi 2002) durchgeführt werden. Wie bereits oben erwähnt, simuliert die expertenbasierte Analyse Nutzerverhalten und Nutzermeinungen. Da diese jedoch tatsächlich nicht bekannt sind, kann das Expertenurteil von dem der wahren Nutzer abweichen.
- **Komparative Analyse von Nutzerverhalten:** Zur komparativen Analyse von Nutzerverhalten auf Basis von Logdaten werden einfache Zugriffsstatistiken (z.B. Hightower u. a. 1998) oder auch komplizierte Nutzungsmuster (z.B. Spiliopoulou u. Berendt 2001) betrachtet. Hierbei können verschiedene Websites oder aber verschiedene Nutzergruppen einer Website miteinander verglichen werden. Ein in der Forschungsliteratur bestehender Ansatz, welcher den Vergleich kompletter Nutzungsmuster zwischen verschiedenen Websites in der in vorliegender Arbeit durchgeföhrten Form beinhaltet, ist dem Autor nicht bekannt.
- **Komparative Analyse von Nutzermeinungen:** Das wohl am häufigsten eingesetzte Instrument zur Analyse von Nutzermeinungen ist der strukturierter Fragebogen. Wenn die Erfassung der Antworten mittels Ratingskalen erfolgt, können u. a. Umfrageergebnisse zu verschiedenen Websites miteinander verglichen werden. Ähnlich zum Ansatz der Multikriteriellen Entscheidungsanalyse errechnen einige Arbeiten hierbei einen Gesamtperformancewert zu

einer Site, allerdings ohne Stakeholderpräferenzen einzubeziehen (z.B. Fornell 1992; Fornell u. a. 1996).

- **Komparative Analyse mit MCDA:** MCDA-basierte Verfahren wurden bereits zur Bewertung von Websites vorgeschlagen. Beispiele hierfür sind die Bewertung auf Basis der Analyse von Nutzermeinungen (z.B. Grigoroudis u. a. 2008; Sampson u. Manouselis 2005) oder mittels Expertenurteil (z.B. Moustakis u. a. 2004).

In keinem der bekannten MCDA-Ansätze zur Bewertung von Websites bzw. Portalen wird mehr als ein Bereich des Systems einbezogen, d.h. es werden *nur* Nutzermeinungen oder *nur* die Ergebnisse einer expertenbasierten Analyse betrachtet. Dementsprechend ist hier auch keine Arbeit bekannt, welche eine Integration verschiedener Bereiche in einem Gesamtmodell vornimmt und daraus eine Gesamtbewertung ableitet.

Gegenstand und Forschungsbeitrag dieser Arbeit ist die Entwicklung eines multiatributiven Entscheidungsmodells zur Bewertung des Erfolgs nicht-kommerzieller Informationsportale. Im Unterschied zum einfachen Benchmarking fließen hierbei die Präferenzen von Stakeholdern mit in den Prozess der komparativen Analyse ein. Außerdem integriert das Modell im Gegensatz zu den bekannten MCDA-Ansätzen zur Bewertung von Websites mehrere Bereiche des zu evaluierenden Systems. Es werden Nutzerverhalten, Nutzermeinungen sowie Aufwendungen zur Pflege der Inhalte betrachtet und eine Gesamtbewertung wird abgeleitet.

1.3 Struktur und Inhalt der Arbeit

Der verbleibende Teil dieser Arbeit ist wie folgt gegliedert:

Kapitel 2 enthält eine kurze Erläuterung zu Webportalen. Die Besonderheiten von Portalen werden charakterisiert und Hauptunterscheidungsmerkmale zu klassischen Websites werden aufgezeigt. Ein Klassifikationsschema anhand der Merkmale (1) Portaltyp, (2) -zweck und (3) -ausrichtung wird vorgeschlagen.

Kapitel 3 enthält eine Einführung zu IS. Die Charakteristika von IS werden beschrieben und die sich daraus ergebende Problematik einer Bewertung wird erläutert. In diesem Kontext werden wichtige Modelle zu relevanten Variablen von IS-Erfolg vorgestellt. Darauf basierend, wird ein Rahmenmodell entwickelt, welches bei der Konzeption eines adäquaten Evaluationsansatzes für ein spezifisches IS in einem spezifischen Kontext Hilfestellung geben soll. Dieses Framework ist Grundlage für die Entwicklung des in dieser Arbeit vorgeschlagenen Bewertungsmodells.

Kapitel 4 gibt eine Einführung zu MCDA. Das Grundmodell der Multikriteriellen Entscheidungsanalyse wird vorgestellt, ein Überblick zu den wichtigsten Verfahren wird gegeben und es wird detailliert auf das in dieser Arbeit verwendete Verfahren – die Nutzwertanalyse – eingegangen.

Kapitel 5 entwickelt das Bewertungsmodell auf Grundlage des Rahmenmodells aus Kapitel 3 und der in Kapitel 4 eingeführten Nutzwertanalyse. Die Bewertungsstrategie zum Einsatz dieses Modells wird dargestellt.

In Kapitel 6 wird das Modell in einer Fallstudie mit vier deutschen Bildungs-server-Portalen angewendet. Eine kurze Einführung zu den betrachteten Portalen wird gegeben. Anschließend werden die Prozesse der Datenbeschaffung, Datenaufbereitung, Ergebnisherleitung und -interpretation zur Bewertung von Nutzerverhalten, Nutzermeinungen und Aufwendungen beschrieben. Die Resultate zu diesen Bereichen werden in einer Gesamtbewertung zusammengefasst. Um den praktischen Nutzen des hier vorgeschlagenen Ansatzes beispielhaft zu belegen, folgt eine Darstellung der aus den Untersuchungsergebnissen gezogenen Schlussfolgerungen zu Verbesserungen für einen der teilnehmenden Portalbetreiber.

Kapitel 7 fasst die wichtigsten Punkte dieser Arbeit zusammen und gibt abschließend einen Ausblick zu möglichen zukünftigen Fragestellungen, die sich aus dem hier vorgestellten Ansatz ergeben.

Kapitel 2

Webportale

Vorliegendes Kapitel gibt eine Einführung zu Webportalen. In Abschnitt 2.1 werden Webportale charakterisiert und Unterschiede zu klassischen Websites aufgezeigt. Abschnitt 2.2 beinhaltet die hier vorgeschlagene Klassifikation von Portalen entsprechend der Kriterien (1) Portaltyp, (2) -zweck und (3) -ausrichtung. Abschnitt 2.3 fasst das Kapitel zusammen.

2.1 Portalkonzept

Ein Webportal ist eine spezielle Form von Website. Der Begriff „Portal“ leitet sich vom lateinischen Begriff *porta* ab, welcher mit „Pforte“, „Tor“ oder „repräsentativ gestalteter Eingang“ übersetzt werden kann. Die Bezeichnung deutet auf den Hauptzweck eines Portals hin: Es dient als Einstiegspunkt zum WWW bzw. zum Intranet einer Organisation. Folgende Definition eines Portals wird hier zugrunde gelegt:

Portal is a term, generally synonymous with gateway, for a World Wide Web site that is or proposes to be a major starting site for users when they get connected to the Web or that users tend to visit as an anchor site. (SearchCIO-Midmarket.com 2003)

Der technische Aufbau eines Portalsystems besteht aus der Präsentationsschicht, der Anwendungslogik und dem Backend. Diese Komponenten setzen auf einer Webserver-Architektur auf (Abbildung 2.1). Die Präsentationsschicht bereitet die Ausgabe für die unterschiedlichen Endgeräte der Nutzer wie z. B. *Hypertext Markup Language* (HTML)-Browser oder Geräte mit *Wireless Application Protocol* (WAP)-Unterstützung auf. Die Anwendungslogik beinhaltet die eigentliche Funktionalität des Portals wie z. B. Content Management oder Rechteverwaltung. Die Backend-Ressourcen bilden die Basis des Systems wie z. B. Datenbanken und Betriebssystem.

Eine der wichtigsten Funktionen von Portalen, welche zugleich ein Hauptunterscheidungsmerkmal zu klassischen Websites darstellt, ist die Erfassung von Informationen aus verschiedenen Quellen, deren Aufbereitung und Präsentation in einer integrierten Umgebung. In der Regel sind den Inhalten Verweise (engl.: *hyperlinks*) zu den Originalquellen und damit zu weiteren Informationsdetails beigegeben. Zusätzlich bieten viele Portale verschiedene Tools und Dienste zur Unterstützung persönlicher Produktivität (z. B. eMail, Adressbuch, Terminkalender), Kommunikation (z. B. Chat, Foren) und Zusammenarbeit (z. B. Groupware) an (Butters 2003; Looney u. Lyman 2000).

Eine eindeutige Trennlinie zwischen klassischer Website und Webportal lässt sich nicht ziehen. Bildlich gesprochen, bewegt sich eine Site ihren Eigenschaften entsprechend auf einem Kontinuum, an dessen einem Ende der Prototyp einer rudimentären Website und an dessen anderem Ende der Prototyp eines hoch komplexen Portals stehen (Miller 2003). Dementsprechend ist jede Website zu einem gewissen Anteil auch ein Webportal und umgekehrt. Historisch gesehen, entwickelte sich das Portalkonzept sehr wahrscheinlich aus Internet-Suchdiensten wie Yahoo (www.yahoo.com) oder Lycos (www.lycos.com). Diese dienten ursprünglich ausschließlich der Suche, also als Tor bzw. Durchgang (engl.: *gateway*) zu anderen Webressourcen. Im Laufe der Zeit wurden dann weitere Funktionen wie z. B. eMail, Kalender oder Chat hinzugefügt, was zu den heute bekannten Formen dieser Sites führte. Sie entwickelten sich somit von reinen Suchagenten zu Plattformen der

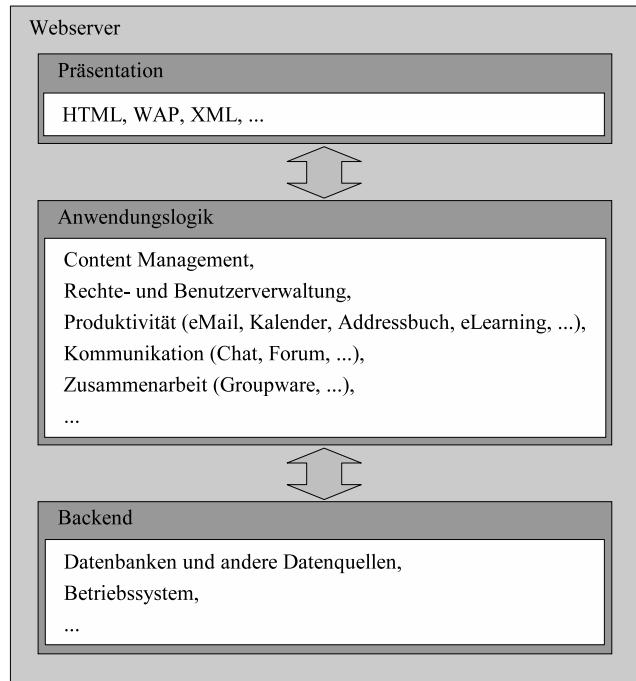


Abbildung 2.1: Schema des technischen Aufbaus eines Webportals

Kommunikation, Zusammenarbeit, Unterhaltung u. a. Damit sind das die ersten Websites, die als Portale bezeichnet werden konnten (Tatnall 2005).

Heute existieren verschiedene Portalformen, die in unterschiedlichen Bereichen eingesetzt werden und dementsprechend von einander differierende Eigenschaften besitzen können. Im folgenden Abschnitt wird dazu ein Klassifikationsansatz vorgestellt.

2.2 Klassifikation von Portalen

Portale können nach Art und Umfang ihrer Eigenschaften in unterschiedliche Klassen gegliedert werden. Entsprechende Vorschläge verschiedener Autoren differieren mehr oder weniger stark von einander (z. B. Looney u. Lyman 2000; Searle 2005; Tatnall 2005; Wojtkowski u. Major 2005).

Mit dem hier vorgestellten Klassifikationsansatz wird eine übersichtliche Darstellung der wesentlichen Kategorien angestrebt, ohne jedoch Anspruch auf Vollständigkeit zu erheben. Die Klassifikation basiert auf der Verwendung von Konzepthierarchien (Wille 1992). Diese ermöglichen die Analyse von Daten auf unterschiedlichen Abstraktions- bzw. Generalisierungsebenen. Eine Konzepthierarchie ist ein Klassifikationsbaum bei dem jeder Knoten genau ein Konzept repräsentiert. Der Baum muss dabei folgende Eigenschaften besitzen:

- Jeder Knoten außer der Wurzel hat genau einen in der Hierarchie darüber

liegenden Knoten.

- Knoten, die zum selben darüber liegenden Knoten gehören, repräsentieren sich gegenseitig ausschließende Konzepte.
- Alle Objekte der zu klassifizierenden Menge können auf den Blattknoten abgebildet werden.

In der Regel ist ein Objekt anhand von nur einer Konzepthierarchie nicht vollständig klassifizierbar. Es können daher mehrere Konzepthierarchien miteinander verknüpft werden. Diese bilden dann einen mehrdimensionalen Merkmalsraum, in welchem das zu klassifizierende Objekt abgebildet wird. Für die hier vorgeschlagene Klassifikation von Portalen werden drei Konzepthierarchien miteinander verknüpft. Diese repräsentieren (1) den Portaltyp, (2) den Zweck des Portals und (3) die Ausrichtung des Portals (Abbildung 2.2).

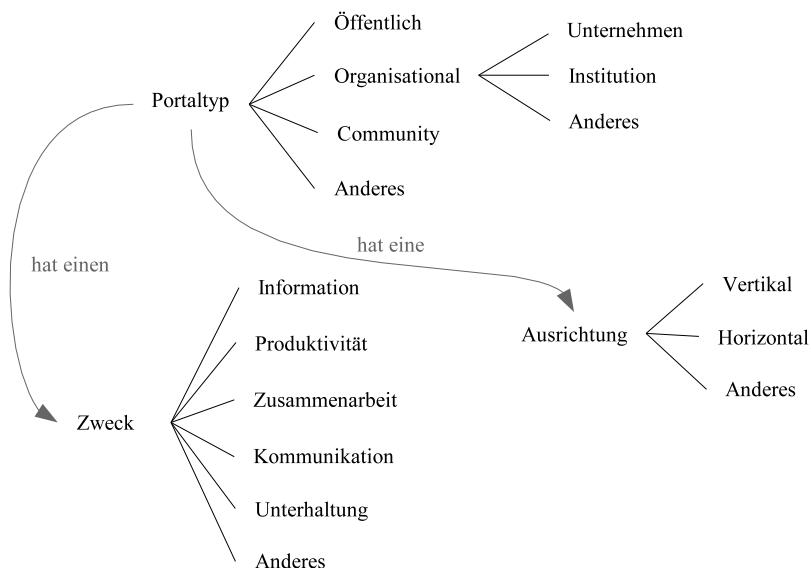


Abbildung 2.2: Konzepthierarchien zur Portalklassifikation

2.2.1 Portaltyp

Portale lassen sich ihrem Typ entsprechend in öffentliche, organisationale und Community-Portale kategorisieren

Öffentliche Portale (engl.: *public portals, general portals*) sind für ein breites Publikum konzipiert. Verschiedene Zielgruppen werden mittels einer Vielfalt an Informationen zu verschiedensten Themen sowie diversen Diensten und Online-Tools angesprochen (Looney u. Lyman 2000; Sieber u. Valor-Sabatier 2005; Tatnall 2005). Beispiel für ein öffentliches Portal ist www.yahoo.de.

Organisationale Portale beinhalten, wie der Name bereits andeutet, spezifische Inhalte zu einer Organisation. Hier kann zwischen Intranet- und Extranetportalen unterschieden werden. Intranetportale sind nur Mitgliedern innerhalb einer Organisation zugänglich. Extranetportale sind auch von außerhalb erreichbar und dienen als Schnittstelle zwischen Organisation und Nicht-Mitgliedern wie z. B. Besuchern, Kunden oder Partnern. Organisationale Portale lassen sich unter anderem in Unternehmensportale und institutionelle Portale aufgliedern.

- Unternehmensportale (engl.: *enterprise portals, corporate portals*) sind Portale von in der Regel kommerziellen Unternehmen, die von Mitarbeitern, Geschäftskunden und -partnern zur Informationsbeschaffung und Unterstützung von Geschäftsprozessen genutzt werden können (Searle 2005; Tatnall 2005). Diese können weiter untergliedert sein in Mitarbeiterportale (engl.: *employee portals*), Endkundenportale (engl.: *consumer portals*), Lieferantenportale (engl.: *supplier portals*) oder Businessportale (engl.: *business portals*) (Kirchhof u. a. 2004). Beispiel für ein Geschäftskunden/-partner bzw. Businessportal ist channel.sap.com.
- Institutionelle Portale sind Portale von staatlichen Einrichtungen, wie z. B. Bildungs- oder Verwaltungseinrichtungen. Diese bieten analog zu den Unternehmensportalen extranet- und intranetspezifische Informationen und Dienste für Angehörige und Besucher der Einrichtung an.

Für Portale höherer Bildungseinrichtungen, wie z. B. Universitäten, wird in der englischsprachigen Literatur häufig auch der Begriff des „institutionellen“ Portals (engl.: *institutional portal*) verwendet (z. B. Bajec 2005; Jacobson 2000; Pearce 2003). Beispiel für ein universitäres Portal ist www.fu-berlin.de.

Portale von z. B. Ämtern und Behörden - so genannte eGovernment-Portale - werden für informative und administrative Zwecke eingesetzt (Aitkenhead 2005). Durch die Nutzung solcher Portale kann man als Bürger zeitaufwändige Behördengänge vermeiden und/oder sich schnell Informationen beschaffen. Beispiel für ein eGovernment-Portal ist www.berlin.de.

Community-Portale sind Plattformen virtueller Gemeinschaften. Eine virtuelle Gemeinschaft bzw. *Community* ist eine Gruppe von Individuen, welche an Stelle des persönlichen Kontakts das Internet als Kommunikations- und Interaktionsmedium nutzen (Li 2004). Auf diese Weise können sich Personen zusammenfinden, die in der realen Welt durch große Distanzen von einander getrennt sind. Community-Portale sind themenspezifische Portale, die basierend auf einem gemeinsamen Interesse von der Onlinegemeinschaft genutzt werden (Looney u. Lyman 2000; Tatnall 2005). Beispiel für ein Community-Portal ist linux-india.org.

2.2.2 Portalzweck

Neben dem Typ kann man auch den Zweck bzw. den Hauptzweck eines Portals als Klassifikationsmerkmal heranziehen. Folgende Merkmale werden betrachtet:

- **Information:** Bereitstellung von Informationen in Form von Texten, Multi-medainhalten und herunterladbaren Dateien.
- **Produktivität:** Bereitstellung von Tools und Diensten, wie z.B. eMail, Adressbuch und Kalender, zur Unterstützung der persönlichen Produktivität.
- **Zusammenarbeit:** Bereitstellung von Tools und Diensten, wie z.B. Groupware, zur Unterstützung der Zusammenarbeit zwischen den Portalnutzern.
- **Kommunikation:** Bereitstellung von Tools und Diensten, wie z.B. Foren, zur Unterstützung der Kommunikation zwischen den Portalnutzern
- **Unterhaltung:** Bereitstellung von Tools und Diensten, wie z.B. Onlinespiele, zu Unterhaltungszwecken.

In der Regel dienen Portale mehr als nur einem Zweck. Dennoch kann man in den meisten Fällen den Hauptzweck bzw. die Hauptzwecke identifizieren. So hat z.B. das oben aufgeführte eGovernment-Portal www.berlin.de den Hauptzweck der Informationsbereitstellung.

2.2.3 Portalausrichtung

Ein weiteres Klassifikationsmerkmal ist die „Ausrichtung“ des Portals. Diese spezifiziert den Grad und die Form der Abdeckung der Themenbereiche durch die auf dem Portal bereitgestellten Informationen. Unterscheiden lassen sich Portale mit vertikaler und horizontaler Ausrichtung.

Vertikale Portale (engl.: *vertical portals*, *vortals*) sind auf einen engen Themenbereich zugeschnitten und beinhalten Informationen und Dienste für ganz bestimmte Nutzergruppen. In der Wirtschaft beispielsweise sind vertikale Portale auf einen speziellen Wirtschaftszweig und damit auf eine bestimmte Gruppe von Unternehmen, die auch entlang einer Wertschöpfungskette angesiedelt sein können, ausgerichtet (Looney u. Lyman 2000; Tatnall 2005). Beispiel für ein vertikales Portal ist das Informationsportal zur stahlerzeugenden und -verarbeitenden Industrie www.stahl-online.de.

Horizontale Portale (engl.: *horizontal portals*) beinhalten eine Vielfalt verschiedener Inhalte und Dienste und sind im Gegensatz zu vertikalen Portalen für eine breite Nutzerschicht mit vielen verschiedenen Interessen konzipiert. Öffentliche Portale sind in der Regel auch horizontale Portale (Sieber u. Valor-Sabatier 2005). Beispiel für ein horizontales Portal ist www.yahoo.de.

2.3 Zusammenfassung

In diesem Kapitel wurden Webportale charakterisiert, Unterschiede zu klassischen Websites hervorgehoben und ein Portalklassifikationsansatz vorgestellt.

Portale sind eine spezielle Form von Websites, die in Abhängigkeit ihres Zwecks u. a. als Plattformen der Informationsgewinnung, Produktivität, Kommunikation

und Zusammenarbeit genutzt werden können. Ein Hauptunterscheidungsmerkmal zu klassischen Websites ist die Integration von Informationen aus verschiedenen Quellen.

Das vorgeschlagene Klassifikationsschema sieht eine Unterscheidung nach Typ, Zweck und Ausrichtung eines Portals vor.

Kapitel 3

Evaluation von Informationssystemen (IS)

Vorliegendes Kapitel gibt eine Einführung zu bestehenden Evaluationsansätzen für IS. Darauf aufbauend werden häufige Ansätze für Websites bzw. Portale vorgestellt.

In Abschnitt 3.1 werden IS charakterisiert und eine kurze Einführung in die Problematik einer Bewertung wird gegeben. Abschnitt 3.2 beinhaltet eine Übersicht relevanter IS-Evaluationsansätze. Darauf basierend wird ein Rahmenmodell entwickelt, welches bei der Konzeption eines adäquaten Bewertungsansatzes für ein spezifisches IS in einem spezifischen Kontext Hilfestellung geben soll. In Abschnitt 3.3 werden verbreitete Ansätze zur Bewertung von Websites vorgestellt und der Bezug zum entwickelten IS-Evaluations-Rahmenmodell wird aufgezeigt. Abschnitt 3.4 fasst das Kapitel zusammen.

3.1 Problematik

Grob gesagt ist ein IS ein computerbasiertes System zur Bereitstellung von Informations- und Kommunikationsdiensten (Davis 1997, S. 138). Welche Eigenschaften ein IS besitzen muss, ist nicht genau festgelegt. Bestehende Systeme können sich in Abhängigkeit ihres speziellen Zwecks voneinander unterscheiden. In der Literatur vorgeschlagene Definitionen variieren daher teilweise. Davis u. Olson (1985) beispielsweise definieren IS als

an integrated, user-machine system for providing information to supreme operations, management, and decision-making functions in an organization. The system utilizes computer hardware and software; manual procedures; models for analysis, planning, control and decision making; and a database. (S. 6)

Sie beziehen sich hierbei auf die eher in Unternehmen vorkommende „traditionelle“, organisationale Form von IS. Diese betrieblichen IS werden auch als *Management Informationssysteme* (MIS) bezeichnet. Eine allgemeinere Definition, die auch eine individuelle, organisationsunabhängige Nutzung einschließt, findet sich auf Wikipedia.org (2008b):

A system, whether automated or manual, that comprises people, machines, and/or methods organized to collect, process, transmit, and disseminate data that represent [...] information.

Ein Webportal ist ein Web-basiertes Informationssystem (WBIS). Diese können sowohl im organisationalen als auch im privaten Umfeld genutzt werden. WBIS fußen auf offenen Standards, wie z.B. *Hypertext Transfer Protocol* (HTTP) und *Hypertext Markup Language* (HTML) (Chandrinos u. Trahanias 1998), und können dadurch kostengünstig betrieben werden. Die Omnipräsenz des Internet erlaubt außerdem eine prinzipiell globale Verfügbarkeit. WBIS ermöglichen damit im Vergleich zu den traditionellen IS günstigere, verbesserte und neue Formen der Geschäftsprozessunterstützung in und zwischen Organisationen (Dunn u. Varano 1999).

Im Folgenden wird in dieser Arbeit sowohl auf Web-basierte als auch auf andere Formen von Informationssystemen mit dem Kürzel „IS“ Bezug genommen. Wenn nötig, wird genauer unterschieden.

Hauptziel des Einsatzes eines IS – unabhängig von dessen konkretem Zweck – ist die Steigerung der Produktivität. Im volkswirtschaftlichen Sinne wird Produktivität als Verhältnis von Input zu Output (z. B. in finanziellen Größen) gemessen. Dieses ist bei einem kommerziellen System leicht umsetzbar. Im nicht-kommerziellen Rahmen ist der Begriff abhängig vom jeweiligen Kontext zu interpretieren. Bei einem nicht-kommerziellen System könnte Produktivität beispielsweise als Verhältnis zwischen eingesetztem Arbeitsaufwand für den Betrieb und dem Ausmaß der Wissensverbesserung seiner Nutzer definiert werden.

Anschaffung und Betrieb eines IS sind mit Kosten verbunden. Da der Einsatz nur dann sinnvoll ist, wenn der Nutzen größer ist als die Kosten, ergibt sich die Notwendigkeit zur Überprüfung der Produktivität. Wie aber kann diese erfasst werden? In der Literatur gibt es dazu viele verschiedene Ansichten und dementsprechend eine Vielzahl unterschiedlicher Ansätze. Allein die Betrachtung der auf die IS-Produktivität möglichen Einflussbereiche lässt auf eine recht große Zahl möglicher Einflussfaktoren schließen. Es werden unter anderem die Umwelt, die Organisation, interorganisationale Beziehungen, die IT-Abteilung, das IT-Design, die IT-Funktion, die IT-Applikation, das IS-Projekt sowie das Individuum (bzw. der Nutzer) genannt (z. B. Larsen 2003; Seddon u. a. 1999; Urbaczewski u. a. 2002). Hinzukommt, dass nicht alle Auswirkungen eines IS direkt erfassbar und/oder finanziell bewertbar sind (Brynjolfson 1993; Chan 2000). Im Rahmen einer Bewertung alle in Frage kommenden Faktoren zu identifizieren und einzubeziehen ist daher praktisch unmöglich.

Die Vielzahl der vorhandenen, unterschiedlichen Methoden, Kriterien und vor allem Meinungen zu diesem Thema kann als Hinweis für eine bestehende Unklarheit bezüglich der Bewertung gedeutet werden. Teile der Forschergemeinde führen diese Unklarheit auf ein ihrer Meinung nach noch fehlendes, gültiges Paradigma der IS-Forschung¹ zurück, welches diese als eigenständige Disziplin legitimieren würde (Keen 1980). Auf der anderen Seite kann die Methodenvielfalt dadurch begründet werden, dass IS auf Grund ihrer vielfältigen Ausformungen und Einsatzmöglichkeiten eine komplexe Welt darstellen, deren Erfassung nicht auf Basis eines einheitlichen Ansatzes durchgeführt werden kann. Zur Diskussion über die Abgrenzung des Forschungsfeldes siehe z.B. Benbasat u. Zmud (2003), Cushing (1990), Keen (1980), Lowry u. a. (2000).

Im folgenden Abschnitt wird eine einführende Übersicht zu vorhandenen Ansätzen gegeben und ein Rahmenmodell entwickelt, welches zur Konzeption eines Evaluationsansatzes für ein konkretes IS herangezogen werden kann.

3.2 Konzeption einer IS-Bewertung

Wie im vorangegangenen Abschnitt angedeutet, resultiert aus dem jetzigen Stand der IS-Forschung eine Vielzahl an Bewertungsansätzen mit unterschiedlichen Methoden und Kriterien. Diese werden in der Literatur meist unter den Oberbegriffen „IS-Produktivität“ (Brynjolfson 1993; Osei-Bryson u. Ko 2004), „IS-Erfolg“ (De-Lone u. McLean 1992, 2003) oder „Usability“ (Hornbæk 2006; Ivory u. Hearst 2001) zusammengefasst.

Ziel der IS-Produktivitätsmessung ist die Erfassung des Verhältnisses von Input in ein IS und des durch das System erzeugten (wirtschaftlichen) Outputs. Gegenstand der IS-Erfolgsbewertung ist die Erfassung und Analyse aller erfolgsrelevanten Eigenschaften eines IS und der Bewertung derer Auswirkungen. Der Begriff „Usa-

¹Das Feld ist im deutschsprachigen Raum Bestandteil des Forschungsfelds der Wirtschaftsinformatik (siehe König 1994).

ability“ kann mit „Bedienbarkeit“ bzw. „Bedienkomfort“ übersetzt werden und lässt sich definieren als „the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use“ (ISO 9241-11 1998).

In der Literatur vorhandene Ansätze, die unter diesen drei Oberbegriffen eine IS-Evaluation durchführen, unterscheiden sich in der Wahl ihrer Bewertungskriterien im Wesentlichen nicht voneinander. Aus dieser recht starken Überlappung ist eine eindeutige Trennung in drei separate Methodenklassen der IS-Evaluation nicht möglich. Unter allen drei Oberbegriffen werden Ansätze betrachtet, die sich mit der Analyse eines oder mehrerer folgender Punkte befassen: Einflussfaktoren auf das System, Systemqualität, Qualität bereitgestellter Informationen und Services, Systemnutzung, Nutzerzufriedenheit und individuelle Auswirkungen. Klar unterscheidbar ist lediglich der Fokus der Bewertung. Im Gegensatz zur Usability-Analyse werden unter „IS-Produktivität“ und „IS-Erfolg“ zusätzlich Ansätze erfasst, die neben individuellen auch organisationale und darüber hinaus gehende Auswirkungen betrachten.

Im Folgenden wird in dieser Arbeit der Begriff „IS-Erfolg“ überwiegend und stellvertretend für „IS-Produktivität“ und „Usability“ verwendet. Wenn nötig werden an entsprechender Stelle die Begriffe „IS-Produktivität“ bzw. „Usability“ explizit genutzt.

Um aus der Vielfalt vorhandener Ansätze die für eine konkrete Bewertungssituation geeigneten auszuwählen, ist das für eine bestimmte Bewertung spezifische Konzept von „Erfolg“ zu klären. Zu diesem Zweck wird ein Rahmenmodell entwickelt, welches eine Erweiterung der von Cameron (1980) vorgeschlagenen „kritischen Fragen“ zur Beurteilung organisationaler Effektivität ist. Dieses soll Hilfestellung bei der Konkretisierung der Evaluation geben. Dargestellt ist es in einem pyramidenförmigen Aufbau, um zu verdeutlichen, dass jede Ebene dieser Pyramide auf der Vorgängerebene aufbaut (Abbildung 3.1). Beginnend mit der untersten, sollte auf jeder Ebene das Evaluationskonzept konkretisiert werden, bevor zur Fragestellung der darüber liegenden Ebene gewechselt wird.

Auf der untersten Ebene werden das zu bewertende IS und der betrachtete Aktivitätsbereich spezifiziert. Darauf aufbauend werden Bewertungsperspektiven, Bewertungszweck, betrachteter Zeitrahmen, einbezogene Analyseebenen, herangezogener Bewertungsmaßstab, angewendete Methoden der Datenerfassung und -analyse und Bewertungskriterien konkretisiert. Der Charakter der Fragestellung zu einer jeweiligen Ebene wandelt sich vom strategischen zum operationalen, je weiter man sich in der Pyramide nach oben bewegt.

Dieses in Abbildung 3.1 dargestellte Rahmenmodell wird im Weiteren als „IS-Evaluationspyramide“ bezeichnet.

3.2.1 Gegenstand und Aktivitätsbereich

IS können abhängig von ihrem jeweiligen Zweck in verschiedenen Bereichen eingesetzt werden und können daher unterschiedliche Ausformungen besitzen. Für die

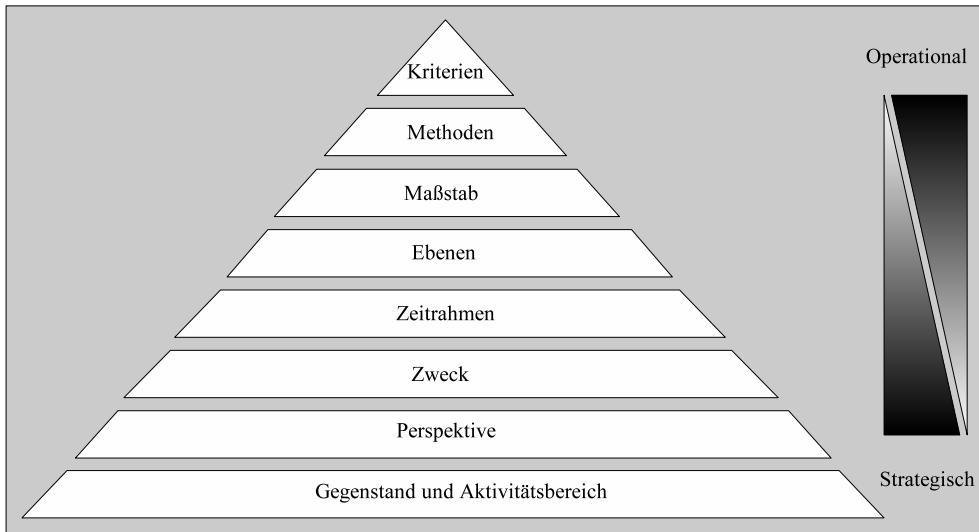


Abbildung 3.1: IS-Evaluationspyramide

Konzeption einer Evaluation muss als erstes spezifiziert werden, was genau der zu evaluierende Gegenstand ist und welcher Aktivitätsbereich (engl.: *domain of activity*) des Systems betrachtet werden soll. Auf dieser Ausgangsbasis baut die weitere IS-spezifische Konzeption des Evaluationsansatzes auf (ähnlich zu den Ansätzen in Cameron 1980; Grover u. a. 1996).

3.2.2 Bewertungsperspektive

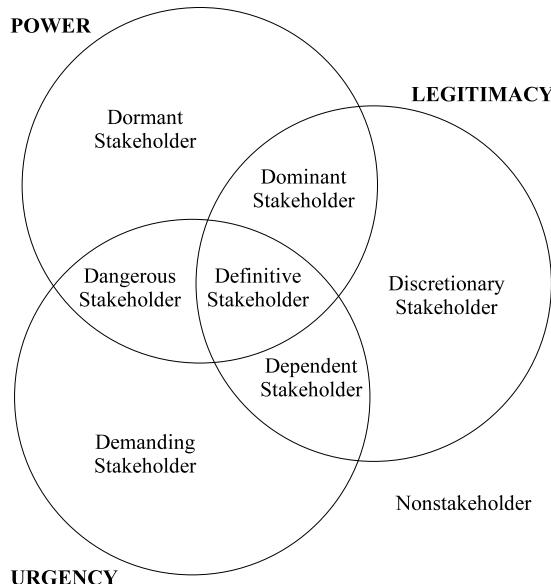
Erfolgsbewertung basiert immer auf den Wertvorstellungen einer oder mehrerer Stakeholdergruppen. Der Begriff „Stakeholder“ lässt sich wie folgt definieren:

A stakeholder in an organization is any group or individual who can affect or is affected by the achievement of the organization's objectives. (Freeman 1984, S. 46)

Erfolg und Erfolgsniveau können von verschiedenen Stakeholdergruppen unterschiedlich definiert und eingeschätzt werden. Möglich sind daher auch konfigurierende Interessen. Um ein vollständiges Bild zu bekommen, müssten daher *alle* Stakeholdergruppen in die Evaluation einbezogen werden. Das ist meist nicht umsetzbar, weil der Stakeholderkreis sonst zu groß würde und damit der Aufwand nicht bewältigbar wäre. Oft ist es auch gar nicht möglich, alle Stakeholder zu identifizieren (z. B. alle Besucher einer Website). In der Regel beschränkt man sich daher auf eine oder wenige relevante Gruppen. Dafür ist festzulegen, welche Stakeholderinteressen vorrangig berücksichtigt werden sollen bzw. aus welchen Stakeholderperspektiven die Bewertung stattfinden soll (siehe Cameron 1980; Grover u. a. 1996; Seddon u. a. 1999). Eine Auswahl kann z. B. anhand der Attribute „Macht und Einfluss“

(engl.: *power*), „Legitimität des Anspruchs“ (engl.: *legitimacy*) und „Dringlichkeit des Anspruchs“ (engl.: *urgency*) vorgenommen werden (Mitchell u. a. 1997), siehe Abbildung 3.2. Diese Attribute werden wie folgt definiert:

- **Power:** „[A] party to a relationship has power, to the extent it has or can gain access to coercive, utilitarian, or normative means, to impose its will in the relationship“ (Mitchell u. a. 1997, S. 865).
- **Legitimacy:** „A generalized perception or assumption that the actions of an entity are desirable, proper, or appropriate within some socially constructed systems of norms, values, beliefs, and definitions“ (Suchmann 1995, S. 547).
- **Urgency:** „The degree to which stakeholder claims call for immediate attention“ (Mitchell u. a. 1997, S. 867).



Quelle: Mitchell u. a. (1997)

Abbildung 3.2: Stakeholdertypologie

3.2.3 Bewertungszweck

In der Evaluationsforschung werden zwei grundlegende Arten des Bewertungszwecks betrachtet: summative und formative Evaluation (Shadish u. a. 1991, S. 78 f.).

Gegenstand einer summativen Evaluation ist die Erfassung, Auswertung und Bewertung von Ergebnissen bzw. Auswirkungen. Die Resultate sollen in erster Linie als Entscheidungsgrundlage für weitere Prozesse dienen.

Ziel einer formativen Evaluation ist die Überprüfung und Korrektur bzw. Verbesserung des Systems. Neben der Analyse auf Korrektheit von z. B. Planung, Implementation, Effektivität und Effizienz kann diese auch die Erforschung der Ursachen für die erzielten Auswirkungen beinhalten. Aus den Resultaten können Verbesserungskonzepte abgeleitet werden. Möglich ist ein sich daran anschließender neuer Evaluationszyklus zur Überprüfung der realisierten Modifikationen und als Basis für weitere Verbesserungskonzepte. Der gesamte Prozess ist in einer sich dauerhaft wiederholenden Form denkbar, mit dem Ziel, den Erfolg des Systems langfristig zu steigern. Auf diese Weise kann auch laufend überprüft werden, ob das System noch aktuelle und neue Anforderungen erfüllt. So können sich beispielsweise Stakeholderbedürfnisse im Laufe der Zeit ändern.

Eine formative Evaluation ist in der Regel umfangreicher und schwieriger durchzuführen als eine summative, da neben der Erfassung von Ergebnissen auch mögliche Ursachen betrachtet und Verbesserungsansätze konzipiert werden.

3.2.4 Zeitrahmen

Langzeiteffekte können sich von Kurzzeiteffekten unterscheiden. Je nach dem welche Ziele der Bewertung zugrunde liegen, ist ein entsprechender Zeitrahmen der Datenerfassung und -auswertung zu wählen (Cameron 1980).

3.2.5 Analyseebenen

Eine Analyse kann auf verschiedenen Ebenen (z. B. Umwelt, Organisation, Gruppe, Individuum oder System) eines IS vorgenommen werden (siehe Cameron 1980; Chan 2000; Grover u. a. 1996; Larsen 2003; Seddon u. a. 1999; Urbaczewski u. a. 2002). Das Erfolgsniveau kann sich hierbei auf den verschiedenen betrachteten Ebenen unterscheiden oder sogar konfligierend sein, wie z. B. hohe Nutzerzufriedenheit gegen geringe Systembetriebskosten (Cameron 1980). Für ein möglichst vollständiges Bild sollten daher mehrere Ebenen berücksichtigt werden.

3.2.6 Bewertungsmaßstab

Mit dem verwendeten Bewertungsmaßstab wird festgelegt, wie die Ergebnisse zu interpretieren sind. In der Evaluationsforschung werden unterschiedliche Bewertungsmaßstäbe vorgeschlagen (Shadish u. a. 1991, S. 78 f.). Beispiele häufiger Ansätze sind die zielbasierte Evaluation (engl.: *goal based evaluation, goal centered evaluation*), die normative Evaluation (engl.: *normative evaluation*), die kognitive Evaluation (engl.: *comparative evaluation*) und die Evaluation der Verbesserung (engl.: *improvement evaluation, reflexive controls*) (Cameron 1980; Grover u. a. 1996; Rossi u. a. 2004, S. 289 ff.).

Bei der zielbasierten Evaluation werden die Ergebnisse anhand gesetzter Ziele bewertet. Diese Art der Bewertung kann ohne weitere, externe Informationen durchgeführt werden. Voraussetzung ist, dass realistische Ziele formuliert werden können.

Bei der normativen Evaluation werden die Ergebnisse anhand von Ideal- oder Standardwerten bewertet. Diese Art der Bewertung ermöglicht eine objektive Beurteilung der Performance. Voraussetzung ist, dass Idealwerte bzw. Standards vorhanden sind. In Abhängigkeit vom Aktivitätsbereich eines IS ist das nicht immer der Fall.

Bei der komparativen Evaluation werden die gemessenen Leistungswerte mit den Werten anderer ähnlicher Systeme verglichen und anhand des Vergleichs bewertet. Diese Art der Bewertung kann ein realistisches Bild der Performance im Vergleich zu den Anderen liefern und hilfreich bei der Identifikation von Schwachpunkten und bester Vorgehensweisen (engl.: *best practices*) sein. Voraussetzung hierfür ist, dass andere ähnliche Systeme zum Vergleich vorhanden sind und in die Evaluation einbezogen werden können.

Bei der Evaluation der Verbesserung wird die Performance anhand von in der Vergangenheit erzielten Ergebnissen bewertet. Diese Art der Bewertung liefert objektive Werte zu erzielten Leistungsunterschieden, erfolgt jedoch abgekapselt von der „Außenwelt“. Es kann somit keine Aussage darüber gemacht werden, wie hoch das absolute Erfolgsniveau des Systems ist (z. B. im Vergleich zu Standards oder anderen Systemen).

3.2.7 Methoden der Datenerfassung und -analyse

In der Literatur werden verschiedene Methoden, wie z. B. die Umfrage oder die Fallstudie, für unterschiedliche Situationen vorgeschlagen (z. B. Urbaczewski u. a. 2002). Diese lassen sich in quantitative und qualitative Ansätze unterteilen (z. B. Chan 2000; Galliers 1992; Hamilton u. Ives 1992), wobei einige, wie z. B. die Fallstudie (Yin 2003), je nach konkreter Ausprägung sowohl quantitativen als auch qualitativen Charakter haben können (Abbildung 3.3).

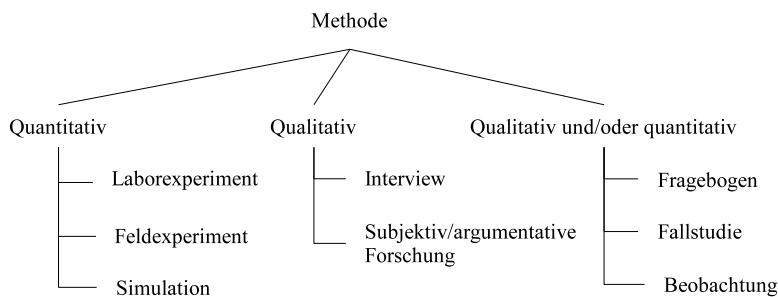


Abbildung 3.3: Aus den betrachteten Ansätzen synthetisierte Klassifikation von Methoden zur Datenerfassung und -analyse

Je nach Gegenstand einer Evaluation und dem zugrunde gelegten Forschungsparadigma können verschiedene Methoden miteinander kombiniert werden. In der Literatur sind drei Paradigmen besonders vertreten: Positivismus, Interpretivismus und Pluralismus.

Der positivistische Ansatz besagt, dass „einzig wahres“ Wissen nur auf Basis von strikter naturwissenschaftlicher (d. h. quantitativer) Methodik durch positive Bestätigung von Theorien erzeugt werden kann (Mill 1865).

Der interpretivistische Ansatz basiert auf dem Standpunkt, dass das Verständnis sozialer Phänomene (die unter anderem Teil eines IS sind) im Gegensatz zum Vorgehen in den Naturwissenschaften prinzipiell andere Methoden erfordert. Sozialwissenschaftliche Forschung kann nach diesem Standpunkt nicht auf reiner Quantifizierung von sozialen Phänomenen basieren, sondern muss eine qualitative Interpretation der Ereignisse aus dem Blickwinkel und auf Basis der Erfahrungen der beteiligten Personen beinhalten (Weber 1967). Die Hauptkritikpunkte am positivistischen Ansatz sind (Galliers 1992): Soziale Phänomene können unter Umständen auf viele verschiedene Arten interpretiert werden, d. h. es gibt im Gegensatz zur naturwissenschaftlichen Methodik nicht nur die eine richtige Erklärung für ein Ereignis; der Evaluator kann ungewollt durch seine Anwesenheit oder durch sein Handeln Einfluss auf das System nehmen; es besteht die Gefahr der sich selbst erfüllenden Prophezeiung, wodurch Objektivität nicht mehr gegeben wäre.

Die traditionelle Sichtweise zum Einsatz unterschiedlicher Methoden sieht für eine Evaluation eine strikte Trennung zwischen positivistischen und interpretivistischen Ansätzen, d. h. zwischen quantitativen und qualitativen Methoden, vor. Hauptargument dafür ist, dass die zugrunde liegenden Paradigmen dieser zwei Methodenklassen auf unterschiedlichen und sich in Teilen widersprechenden Annahmen basieren. In pluralistischen Ansätzen wird dieser Ansicht widersprochen und eine Kombination befürwortet. Dort liegt die Annahme zugrunde, dass die Kombination verschiedener Methodenklassen eine Möglichkeit zur Verbesserung einer Evaluation ist. Auf diese Weise können mehr Kriterien berücksichtigt und das Evaluationsobjekt aus vielen verschiedenen Perspektiven betrachtet werden, was letztendlich zu reichhaltigeren und damit nützlicheren Evaluationsergebnissen führt (Mingers 2001).

Quantitative Methoden

Quantitative Methoden sind durch eine hypothetisch-deduktive Vorgehensweise charakterisiert, d. h. Hypothesen werden aufgestellt und mit Hilfe statistischer Analysemethoden unter möglichst kontrollierten (laborartigen) Bedingungen überprüft. Sie werden daher auch als „naturwissenschaftliche“ Methoden bezeichnet (Galliers 1992, S. 147 f.) und basieren auf der Annahme, dass systematische und stabile Beziehungen zwischen Phänomenen des Evaluationsobjektes vorliegen und objektiv mittels eines Modells der Realität erfasst werden können. Ihre Hauptcharakteristiken sind eine vereinfachte Abbildung der Realität in einem Modell und die Wiederholbarkeit der untersuchten Abläufe. Als Vorteile quantitativer Methoden werden die Objektivität und damit Glaubwürdigkeit von Ergebnissen sowie deren Vergleichbarkeit und Verallgemeinbarkeit angesehen. Probleme können bei der Analyse komplexer (z.B. sozialer) Systeme auftreten. Auf solche Systeme können viele verschiedene Faktoren, die unter Umständen nicht alle quantitativ erfassbar

sind, direkt oder indirekt einwirken. Eine Abbildung der erfassbaren Zusammenhänge im Modell wäre dann möglicherweise eine zu starke Vereinfachung der Realität. Dadurch würde es schwierig im realen System auftretende Phänomene auf den Einfluss von nur ganz bestimmten Faktoren zurückzuführen und es bestünde das Risiko, dass das Modell eine verfälschte Darstellung der Zusammenhänge lieferte.

Zu den quantitativen Ansätzen gehören unter anderem das Laborexperiment, das Feldexperiment und die Simulation (Galliers 1992).

Laborexperiment/-studie Beim Laborexperiment wird eine präzise Analyse des Einflusses der Variation einer oder mehrerer unabhängiger Variablen auf die Ausprägungen einer oder mehrerer abhängiger Variablen innerhalb einer künstlichen, kontrollierten Umgebung vorgenommen. Die begrenzte Zahl an betrachteten Variablen erlaubt ein genaues Studium von Zusammenhängen. Die künstliche Umgebung bildet jedoch eine sehr starke Vereinfachung der Realität. Die Gefahr ist daher groß, dass wichtige Faktoren unberücksichtigt bleiben. Unter Umständen kann diese Umgebung die Zusammenhänge zwischen den untersuchten Variablen sogar beeinflussen.

Ein Beispiel für die Verwendung einer entsprechenden Methode bei der Analyse von IS findet sich in Benbunan-Fich (2001).

Feldexperiment/-studie Beim Feldexperiment wird die Studie aus dem Labor in eine Realweltumgebung verlagert. Zusammenhänge können so unter realistischen Bedingungen analysiert werden. Die Umgebung und deren Einfluss auf das Evaluationsobjekt kann bei dieser Methode allerdings nur in geringem Umfang oder gar nicht kontrolliert werden.

Ein Beispiel für die Verwendung einer entsprechenden Methode bei der Analyse von IS findet sich in Drèze u. Zufryden (1997).

Simulation Bei der Simulation wird das Verhalten eines Systems mit zufällig erzeugten Werten kopiert. Auf diese Weise können Situationen erzeugt und betrachtet werden, welche in der realen Welt nur unter erschwerten Bedingungen oder gar nicht analysierbar wären.

Eine andere Form der Simulation ist das Rollenspiel. Bei diesem bedient und bewertet beispielsweise ein Evaluator das System aus der Rolle eines Nutzers. Nachteile (wie beim Experiment) ergeben sich auch hier insbesondere aus der Vereinfachung der Realität.

Ein Beispiel für die Verwendung einer entsprechenden Methode bei der Analyse von IS findet sich in Palmer (2002).

Qualitative Methoden

Aus den Schwierigkeiten heraus, mit der „naturwissenschaftlichen“ Vorgehensweise soziale Phänomene zu erforschen, wurden Methoden entwickelt, deren Wurzeln in der Anthropologie liegen. Diese basieren auf der Erfassung qualitativer Daten durch

z. B. Beobachtung von Strukturen und Abläufen oder durch intensive Kommunikation mit Beteiligten (siehe Shadish u. a. 1991, S. 270 ff.). Gegenstand qualitativer Methoden ist die Erfassung des Evaluationsobjekts und seiner Auswirkungen aus verschiedenen Perspektiven und in Zusammenhang mit den sozialen Strukturen, die es umgeben. Ansätze die auf diesem interpretivistischen Paradigma basieren, können zu einem besseren Verständnis von Kontextfaktoren und deren Einfluss beitragen. Nachteilig kann sich allerdings der Mangel an Objektivität auswirken, da subjektive Eindrücke die Hauptrolle spielen. Außerdem ist eine Vergleichbarkeit von Untersuchungsergebnissen durch deren Nicht-Quantifizierbarkeit schwierig.

Qualitative Ansätze sind unter anderem das Interview und die subjektiv/argumentative Forschung.

Interview Das Interview kann eine strukturierte, semistrukturierte oder unstrukturierte Form haben. Ein strukturiertes Interview ähnelt in seiner Form einem strukturierten Fragebogen (s. unten) und kann mitunter schon einen quantitativen Charakter haben. Beim semistrukturierten Interview möchte der Befrager zwar Informationen zu bestimmtem Punkten erhalten, die Befragung hat aber mehr den Charakter eines Gesprächs. Beim unstrukturierten Interview sind keine oder nur wenige Befragungspunkte vorgegeben. Der Befragte agiert von sich aus.

Die Befragung gut informierter Personen kann wichtige Einblicke zu der untersuchten Situation liefern. Daneben können zusätzlich wichtige Kontextinformationen erfragt werden. Diese basieren allerdings auf der Sichtweise und Interpretation der befragten Person und können daher subjektiv beeinflusst sein. Außerdem übt der Interviewer möglicherweise durch seine Präsenz und die Art der Befragung Einfluss auf den Befragten aus (Yin 2003, S. 89 ff.).

Ein Beispiel für den Einsatz von Interviews bei der Analyse von IS findet sich in Cutler u. Sterne (2000).

Subjektiv/argumentative Forschung Subjektiv/argumentative Forschung erfolgt auf einer konzeptionellen Ebene, als eine Art individueller, kreativer Prozess. Dieser basiert auf subjektiver Meinung und Spekulation und weniger auf beobachteten bzw. beobachtbaren Fakten. Diese Art der „Forschung“ kann hilfreich für die Bildung von Theorien sein, die anschließend dann auf eine mehr formelle Weise geprüft werden können. Die Methode ist vorteilhaft für die Entwicklung neuer Ideen und Konzepte, jedoch ist der gesamte Prozess unstrukturiert und subjektiv (Galliers 1992).

Ein Beispiel für die konzeptionelle Entwicklung eines Analyse-Modells für IS findet sich in Appan u. Mellarkod (2003).

Methoden mit quantitativem als auch qualitativem Charakter

Es gibt Methoden, die je nach ihrer konkreten Ausprägung sowohl zur quantitativen als auch zur qualitativen Datenerfassung und -analyse verwendet werden können.

Unter anderem sind das die Umfrage per Fragebogen, die Beobachtung und die Fallstudie.

Fragebogen Fragebögen können ähnlich zu Interviews strukturiert oder semi-strukturiert sein. Ein strukturierter Fragebogen hat zu jeder Frage vorgegebene Antwortkategorien, wie z. B. eine Skala von eins bis fünf. So ein Fragebogentyp eignet sich zur quantitativen Datenerfassung. Semistrukturierte Fragebögen beinhalten keine vorgegebenen Antwortkategorien. Der Umfrageteilnehmer formuliert seine Antworten frei. Fragebögen dieses Typs sind für eine qualitative Analyse geeignet. Fragebögen mit Fragen beider Typen haben somit quantitativen als auch qualitativen Charakter.

Der Einsatz eines Fragebogens ermöglicht es, in relativ kurzer Zeit eine große Menge an Individuen zu befragen. Bei strukturierten Fragebögen kann die Datenauswertung zudem recht schnell und einfach durchgeführt werden. Ein Nachteil ist die Unsicherheit bezüglich des Verständnisses der Fragen. Ein Proband hat hier nicht die Möglichkeit, wie z. B. bei einem Interview, Verständnisfragen zu stellen. Des Weiteren kann ein Fragebogen nur eine beschränkte Menge von Informationen aufnehmen. Die Form der Fragestellungen kann außerdem Einfluss auf die Beantwortung haben.

Eine Umfrage per Fragebogen erfasst in der Regel nur eine Teilmenge der Grundgesamtheit aller anvisierten Zielpersonen. Je nach Auswahl der Stichprobe können die erzielten Ergebnisse ein verzerrtes oder gar falsches Bild der Grundgesamtheit wiedergeben.

Ein häufig auftretendes Problem ist außerdem die Nicht-Beantwortung von Fragen. Unvollständig beantwortete Fragebögen sind für eine Analyse nur beschränkt oder gar nicht brauchbar. Um auch diese in die Auswertung einzubeziehen, können fehlende Werte ergänzt werden. Einen einführenden Überblick zu entsprechenden Methoden geben Rässler u. Riphahn (2006).

Beispiele für den Einsatz von Fragebögen zur Analyse von IS finden sich in Zviran u. Erlich (2003).

Beobachtung Gegenstand einer Beobachtung können unter anderem Strukturen, Prozesse und Personen sein, die Teil des IS sind oder mit diesem in Zusammenhang stehen. Der Detaillierungsgrad der erfassten Daten kann hierbei von sehr allgemeinen, qualitativen bis hin zu sehr spezifischen, quantitativen Daten reichen.

Mittels Beobachtung können Geschehnisse sowohl in Echtzeit in ihrer natürlichen Umgebung betrachtet werden als auch in bestimmten Datenbeständen (z. B. Logdaten) festgehaltene Ereignisse aus der Vergangenheit. Im Gegensatz zum Experiment findet keine (zumindest absichtliche) Manipulation des Evaluationsobjektes statt. In diesem Rahmen können auch zusätzliche und möglicherweise wichtige Kontextinformationen mit erfasst und in die Analyse einbezogen werden. Unter Umständen erfasst ein Beobachter allerdings nur einen kleinen Teil der Geschehnisse innerhalb des oft kurzen Beobachtungszeitraums. Daraus lässt sich dann nur

ein unvollständiges Bild der betrachteten Situation ableiten. Außerdem beeinflusst unter Umständen die persönliche Sichtweise des Beobachters die Ergebniserfassung und -interpretation (Yin 2003, S. 92 ff.).

Ein Beispiel für die Beobachtung von Systemeigenschaften zur Analyse von IS findet sich in Moustakis u. a. (2004).

Fallstudie Hauptcharakteristikum der Fallstudie ist der Fokus auf die Erfassung reeller Zusammenhänge innerhalb einer Organisation oder einer Gruppe von Organisationen – also eines konkreten Falls bzw. einer Gruppe von konkreten Fällen. Die Fallstudie ist im eigentlichen Sinne keine Methode für sich. Vielmehr können im Rahmen einer Fallstudie dem zugrunde gelegten Evaluationsansatz entsprechend verschiedene Methoden, wie z. B. eine Beobachtung mit einer Umfrage, kombiniert werden (siehe Irani 2002; Paré 2004; Yin 2003).

Die Fallstudie ermöglicht eine sehr realitätsnahe Erfassung von Zusammenhängen. Der Umfang an verfügbaren Analysedaten ist dabei durch die Kombination verschiedener Methoden größer, als beim Einsatz von nur einer Methode. Dies erlaubt eine Betrachtung des Evaluationsobjekts aus verschiedenen Perspektiven (siehe Abschnitt 3.2.2) und auf verschiedenen Ebenen (siehe Abschnitt 3.2.5). Durch die Beschränkung auf eine Organisation bzw. eine Gruppe von Organisationen ist es allerdings schwieriger, allgemein gültige Schlüsse zu ziehen. Je nachdem welche Methoden angewendet werden, ist eine Fallstudie quantitativ, qualitativ oder beides.

Ein Beispiel für den Einsatz einer Fallstudie zur Analyse von IS findet sich in Irani (2002).

3.2.8 Bewertungskriterien

Je nach gewählter Untersuchungsmethode können verschiedene Bewertungskriterien herangezogen werden. Die Vielfalt an vorgeschlagenen Kriterien ist entsprechend der Menge bestehender Ansatzmöglichkeiten groß (siehe z. B. DeLone u. McLean 1992; Grover u. a. 1996; Seddon u. a. 1999). Die Relevanz eines Kriteriums sowie dessen Einfluss auf andere Kriterien werden hierbei in Abhängigkeit des zugrunde gelegten Modells relevanter Variablen von IS-Erfolg erfasst. Je nach Modell (drei wichtige Modelle werden weiter unten im Detail beschrieben) können betrachtete Kriterien in externe Faktoren, modellinterne unabhängige Variablen und modellinterne abhängige Variablen klassifiziert werden.

Externe Faktoren sind tendenziell Einflussfaktoren auf ein IS und dessen Wirkungsweise. Wie bereits in Abschnitt 3.1 angedeutet, kann die Anzahl möglicher externer Einflussfaktoren sehr groß sein und reicht von Umwelteigenschaften (z. B. das politische System einer Gesellschaft) bis hin zu individuellen Eigenschaften des Nutzers (z. B. persönliche Einstellungen und Neigungen), siehe Larsen (2003). Entsprechende Faktoren als Bewertungskriterien zur Erfassung von IS-Erfolg zu verwenden, wäre nicht sinnvoll, da diese vom System nicht beeinflusst werden können. Wohl aber sind sie im Rahmen einer konkreten Evaluation als Quellen relevanter

Kontextinformationen zu berücksichtigen, und die Ergebnisse sind entsprechend zu interpretieren.

Auf Basis einer umfangreichen empirischen Analyse angewandter Kriterien zur IS-Bewertung identifizieren DeLone u. McLean (1992, 2003) sechs wichtige Kategorien bzw. Variablen von IS-Erfolg: (1) Informationsqualität, (2) Systemqualität, (3) Servicequalität, (4) Systemnutzung, (5) Nutzerzufriedenheit und (6) Systemauswirkungen (siehe Abbildung 3.4). Diese Kategorien werden im Weiteren detailliert beschrieben. Der Begriff „Qualität“ spielt hierbei eine zentrale Rolle und wird nach DIN ISO Norm 8402 definiert als

die Gesamtheit von Eigenschaften und Merkmalen eines Produktes oder einer Dienstleistung, die sich auf deren Eignung zur Erfüllung festgelegter oder vorausgesetzter Erfordernisse beziehen.

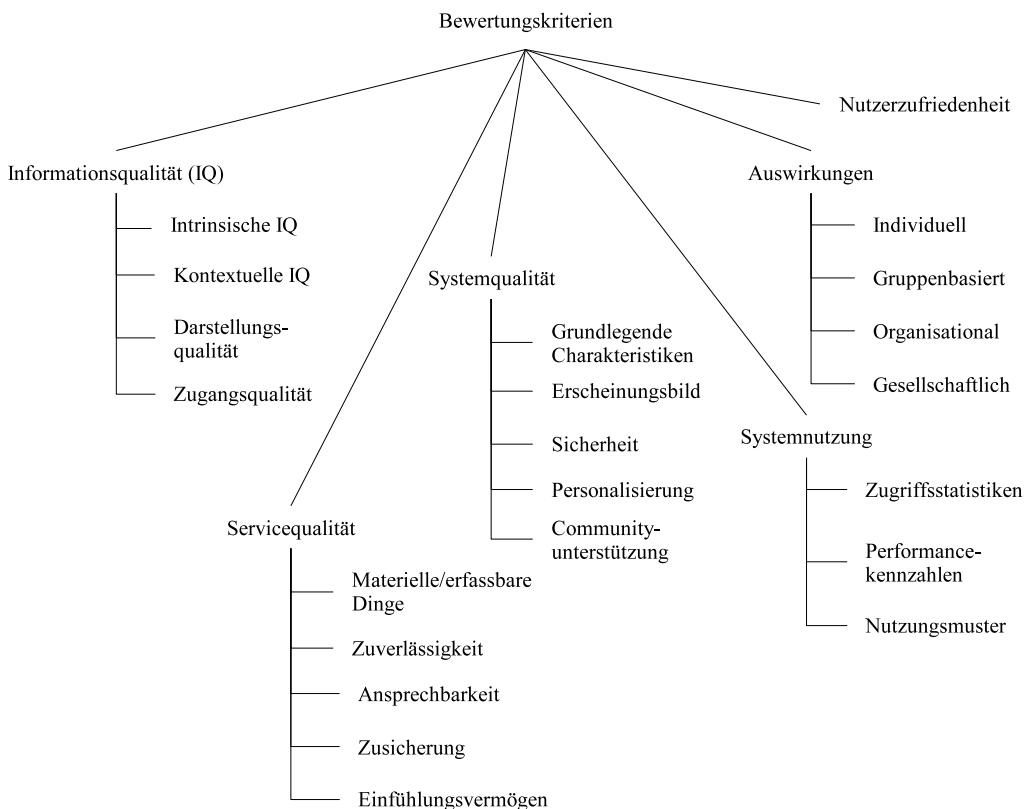


Abbildung 3.4: Aus den betrachteten Ansätzen synthetisierte Klassifikation von IS-Bewertungskriterien

Informationsqualität

Hauptzweck eines IS ist die Verarbeitung und Bereitstellung von Informationen. Informationsqualität (IQ) bzw. die Qualität der zugrunde liegenden Daten (Albrecht 2001) ist somit ein zentraler Punkt für den Erfolg des Systems. Bewertungskriterien bzw. Qualitätsmerkmale für Informationsqualität lassen sich in vier Klassen aufteilen (Wang u. Strong 1996): (1) Intrinsische IQ (engl.: *intrinsic IQ*), (2) Kontextuelle IQ (engl.: *contextual IQ*), (3) Darstellungs- bzw. Repräsentationsqualität (engl.: *representational IQ*) und (4) Zugangsqualität (engl.: *accessibility IQ*).

Mit „intrinsisch“ bezeichnet man „eine Eigenschaft, die ein Gegenstand, Lebewesen oder System von sich aus aufweist[...]. Sie wohnt dem System inne, und liegt ohne äußere Ursache oder Beeinflussung vor“ (Wiktionary.org 2008). Intrinsische IQ bezieht sich somit auf die Qualität der Information bzw. Daten an sich. Verwendete Bewertungskriterien für intrinsische IQ sind unter anderem: Genauigkeit/Richtigkeit, Glaubwürdigkeit, Objektivität, Reputation, Vertrauenswürdigkeit, Konsistenz und Zuverlässigkeit (Lee u. a. 2002). Kontextuelle IQ bezieht sich auf den Wert der Informationen bzw. Daten im Kontext des Umfeldes und der aktuell durchzuführenden Aufgabe des Systems. Verwendete Bewertungskriterien sind unter anderem: Relevanz, Vollständigkeit, Aktualität, Menge, Detailierungsgrad und Nutzen (Lee u. a. 2002). Darstellungs- bzw. Repräsentationsqualität bezieht sich auf die Informations- bzw. Datenrepräsentation im Rahmen des IS. Verwendete Bewertungskriterien sind unter anderem: Lesbarkeit, Verständlichkeit, Interpretierbarkeit, Form und Exklusivität (Lee u. a. 2002). Zugangsqualität bezieht sich auf die Abrufbarkeit der Informationen bzw. Daten. Verwendete Bewertungskriterien sind unter anderem: Zugänglichkeit, Sicherheit und Lokalisierbarkeit (Lee u. a. 2002).

In der Literatur finden sich zahlreiche Arbeiten, die entsprechende Kriterien zur Analyse von IS (siehe DeLone u. McLean 1992) bzw. WBIS (z. B. Agarwal u. Venkatesh 2002; Alexander u. Tate 1999; Benbunan-Fich 2001; Drèze u. Zufryden 1997; Grandon u. Ranganathan 2001; Knight u. Burn 2005; Moustakis u. a. 2004; Olsina u. a. 1999; Palmer 2002) verwenden.

Systemqualität

Neben der Qualität der präsentierten Informationen wird die Kapazität des Systems zur Informationsverarbeitung und -darstellung, also die Systemqualität, als weitere wichtige Variable von IS-Erfolg erachtet. Ein Einblick zu in der Literatur verwendeten Kriterien bei „traditionellen“ IS findet sich in DeLone u. McLean (1992). Für WBIS sei auf Aladwani u. Palvia (2002) verwiesen. Prinzipiell unterscheiden sich die für unterschiedliche IS bzw. WBIS verwendete Kriterien nicht voneinander. Es gibt sehr wenige Ausnahmen, wie z. B. Communityunterstützung, die eher WBIS-spezifisch sind. Die bestehenden Kriterien lassen sich in fünf Kategorien untergliedern: (1) Grundeigenschaften, (2) Erscheinungsbild, (3) Sicherheit, (4) Personalisierung und (5) Communityunterstützung.

Grundeigenschaften Diese beziehen sich auf nicht-funktionale Eigenschaften des Systems, die notwendig und essenziell für den Betrieb sind. Kriterien, die dieser Kategorie zugeordnet werden können, sind unter anderem: Verfügbarkeit, Zugänglichkeit, Stabilität bzw. Zuverlässigkeit, Interaktivität, Reaktionszeit, Bedienbarkeit/Navigierbarkeit. In der Literatur finden sich zahlreiche Arbeiten, die entsprechende Kriterien für die Analyse von IS (z. B. Bailey u. Pearson 1983; Chin u. a. 1988; Davis 1986, 1989; DeLone u. McLean 1992; Doll u. Torkzadeh 1988; Etezadi-Amoli u. Farhoomand 1996) bzw. WBIS (z. B. Agarwal u. Venkatesh 2002; Aladwani u. Palvia 2002; Benbunan-Fich 2001; Galletta u. a. 2004; Grandon u. Ranganathan 2001; Liu u. Arnett 2000; Moustakis u. a. 2004; Negash u. a. 2003; Olsina u. a. 1999; Palmer 2002; Pardue u. Landry 2001; Sampson u. Manouselis 2005; Yoo u. Donthu 2001; Zhang u. von Dran 2001) verwenden.

Erscheinungsbild Betrachtet wird das Erscheinungsbild (engl.: *look and feel*) der Nutzerschnittstelle (engl.: *user interface*). Verwendete Kriterien sind unter anderem: Visuelle Attraktivität und Konsistenz von Formen, Farben und Layout; Bildschirmschärfe und -helligkeit; Fontgröße und -form; Lesbarkeit der Inhalte; Organisation der Inhalte. Entsprechende Ansätze finden sich für traditionelle IS z. B. in Chin u. a. (1988); Lewis (1993) sowie für WBIS z. B. in Agarwal u. Venkatesh (2002); Aladwani u. Palvia (2002); Drèze u. Zufryden (1997); van der Heijden (2003); Hong u. a. (2004); Moustakis u. a. (2004); Olsina u. a. (1999); Sampson u. Manouselis (2005); Yoo u. Donthu (2001); Zhang u. von Dran (2001).

Sicherheit Der Sicherheitsaspekt ist ein wesentlicher Einflussfaktor auf den Umfang und die Art der Nutzung eines Systems. Dieses gilt insbesondere für WBIS im Rahmen einer freiwilligen Nutzung. Die unpersönliche Natur des Internet, dessen technologische Unvorhersehbarkeit sowie die Ungewissheit bezüglich des Verhaltens von Websiteprovidern machen es zu einem unsicheren Medium für den Nutzer (Bhimani 1996). Dementsprechend werden Sicherheit und Schutz der Privatsphäre in der Regel als wichtig angesehen.

Fundamentale Anforderungen an die Sicherheit entsprechender Systeme lassen sich anhand von fünf Kriterien definieren (Bhimani 1996):

- **Vertraulichkeit (engl.: *confidentiality*)**: Die Kommunikation zwischen zwei Parteien darf nicht von dritten einsehbar sein. Dieser Aspekt ist ein wesentlicher Bestandteil des Schutzes der Privatsphäre der Nutzer (engl.: *user privacy*). Vertraulichkeit wird z. B. durch sichere Datenübertragung wie SSL-Verschlüsselung erzielt.
- **Authentifizierung (engl.: *authentication*)**: Eine Partei soll sich in der „Echtheit“ der gegenüberstehenden Partei sicher sein können. Authentizität wird z. B. durch digitale Signaturen erzielt.
- **Datenintegrität (engl.: *data integrity*)**: Zwischen zwei Parteien fließende Datenströme dürfen nicht von dritten modifizierbar sein. Das gleiche gilt

für gespeicherte Daten. Datenintegrität wird z. B. durch sichere Datenübertragung erzielt.

- **Nicht-Zurückweisung (engl.: *nonrepudiation*):** Eine Partei soll sicher gehen können, dass eine bereits durchgeführte Transaktion nicht im Nachhinein durch die gegenüberliegende Partei zurückgewiesen bzw. nicht-anerkannt wird. Beispiel: Nach Bestellung und Lieferung eines Produktes, wird dieses vom Empfänger unberechtigt zurückgewiesen und eine Zahlung erfolgt nicht. Hier kann z. B. eine digitale Signatur als Beweis sicherstellen, dass ein entsprechender Bestellauflauf eingegangen ist. Auf der anderen Seite können Zertifikate und Gütesiegel eines Anbieters den Eindruck von Seriosität beim Nutzer schaffen bzw. verstärken.
- **Selektive Anwendung von Diensten (engl.: *selective application of services*):** Eine Partei braucht der gegenüberliegenden Partei nicht alle privaten Informationen zur Durchführung einer Transaktion zu geben. Wenn z. B. ein Auftrag per Kreditkarte bezahlt wird, kann der Verkäufer die Kreditkartennummer in verschlüsselter Form erhalten. Diese wird an die entsprechende Bank weitergeleitet und dort entschlüsselt. Auf diese Weise ist der Käufer vor Missbrauch seiner Kreditkartennummer geschützt.

Zu Ansätzen für traditionelle IS siehe z. B. Etezadi-Amoli u. Farhoomand (1996). Zu WBIS-Ansätzen siehe z. B. Gehrke u. Turban (1999), Lala u. a. (2002), Liu u. Arnett (2000), Song u. Zahedi (2001), Yoo u. Donthu (2001), Zhang u. von Dran (2001).

Personalisierung Personalisierung beinhaltet die individuelle Anpassung von Inhalten, Funktionen und Eigenschaften des Systems an nutzerspezifische Charakteristiken (z. B. Admin oder User) und Bedürfnisse (z. B. Layout, Farben, Interessen). Personalisierung ermöglicht dem Nutzer produktiveres Arbeiten und wird insbesondere bei WBIS als eine wichtige Systemeigenschaft betrachtet (siehe Jacobson 2000; Pearce u. a. 2003, S. 42 f.) und als Bewertungskriterium verwendet (z. B. Agarwal u. Venkatesh 2002; Aladwani u. Palvia 2002; Palmer 2002; Sampson u. Manouselis 2005).

Communityunterstützung Virtuelle Communities haben sich zu einem bedeutenden Bestandteil des Internet entwickelt. Wie bereits in Abschnitt 2.2.1 beschrieben, bestehen diese aus Gruppen von Individuen, die an Stelle des persönlichen Kontakts das Internet als Kommunikations- und Interaktionsmedium nutzen (siehe Gupta u. Kim 2004; Li 2004). Mitglieder einer Online-Community können miteinander kommunizieren und interagieren, unabhängig von den in der realen Welt zwischen ihnen liegenden Distanzen. Virtuelle Communities können sich außerdem positiv auf den Erfolg der Nutzung eines IS bzw. WBIS auswirken: In sehr kurzer Zeit können distanzunabhängig Informationen, Daten und Wissen ausgetauscht werden; Nutzer können bei Fragen sehr schnell Hilfe aus der Gemeinschaft

erhalten; die Interaktion in der Gemeinschaft und die Reputation der anderen Mitglieder kann das Vertrauen eines Nutzers in das System stärken (Gupta u. Kim 2004). Communityunterstützung wird damit als ein weiteres wichtiges Kriterium für den Erfolg des Systems gesehen, insbesondere bei WBIS (z.B. Agarwal u. Venkatesh 2002; Gallaugher u. Downing 2005; Sampson u. Manouselis 2005).

Servicequalität

Aus der Marketingforschung ist bekannt, dass Servicequalität ein wichtiger Unternehmenserfolgsfaktor ist (Zeithaml u.a. 1996). Ein IS besteht nicht nur aus einer Software, sondern beinhaltet insbesondere bei WBIS zunehmend auch Servicekomponenten. Diese unterstützen den Nutzer in seiner Arbeit und können somit seine Produktivität steigern. Ein simples Beispiel hierfür ist eine Kundenhotline für Fragen zu technischen Problemen. Entsprechend ist neben der Qualität der Informationsverarbeitung auch die Qualität der Servicekomponente eines IS bzw. WBIS als wichtiger Erfolgsfaktor einzubeziehen (siehe DeLone u. McLean 2003; Liu u. Arnett 2000; Pitt u.a. 1995; van Riel u.a. 2001; Watson u.a. 1998).

Das ursprünglich aus dem Marketing kommende SERVQUAL-Instrument (Parasuraman u.a. 1988) wird hierfür in angepasster Form zur Erfassung von IS (z.B. Kettinger u. Lee 1997; Pitt u.a. 1995) bzw. WBIS-Servicequalität (z.B. Wang u. Tang 2003; Webb u. Webb 2002) vorgeschlagen. Bei diesem Ansatz wird Servicequalität als Unterschied bzw. Lücke (engl.: gap) zwischen empfundener und erwarteter Dienstleistungsqualität definiert und mittels fünf Kategorien erfasst: (1) Materielle, erfassbare Dinge (engl.: tangibles), (2) Zuverlässigkeit (engl.: reliability), (3) Ansprechbarkeit (engl.: responsiveness), (4) Zusicherung (engl.: assurance) und (5) Einfühlungsvermögen (engl.: empathy).

Andere Marketingforscher kritisieren, dass die lückenbasierte Messung von SERVQUAL keine korrekte Erfassung von Servicequalität ermögliche und schlagen eine einfache, Performance orientierte Messung (SERVPERF) auf Grundlage der gleichen Kriterien vor (Cronin u. Taylor 1992). Entsprechende Ansätze werden auch für die Bewertung von IS- (z.B. Landrum u. Prybutok 2004; Van Dyke u.a. 1995) bzw. WBIS-Servicequalität (z.B. Cao u.a. 2004; Gefen 2002b; Liu u. Arnett 2000; Negash u.a. 2003; Palmer 2002) verwendet.

Die Eignung des SERVQUAL- bzw. SERVPERF-Instruments für die Analyse von IS- bzw. WBIS-Servicequalität ist allerdings nicht eindeutig geklärt. Aus den in der Literatur verwendeten Kriterien ergeben sich zum Teil Überlappungen mit Kriterien zur Analyse der Informations- und Systemqualität (siehe Zeithaml u.a. 2002). Es ist unbestritten, dass es eine Servicekomponente gibt, die sich von den anderen IS-Bewertungskategorien unterscheidet und daher separat betrachtet werden muss. Eine klare Abgrenzung als eigenständige IS-Bewertungskategorie liegt aber noch nicht vor (siehe Diskussion in Zeithaml u.a. 2002).

Nutzerzufriedenheit

Zufriedenheit kann im Kontext der Nutzung bzw. des „Konsums“ von Diensten definiert werden als

the consumer's sense that consumption provides outcomes against a standard of pleasure versus displeasure. (Oliver 1999, S. 34)

Als „*standard of pleasure*“ kann hierbei ein aus der Perspektive des Konsumenten und auf Basis vorhandener Ansprüche individuell festgesetzter Standard verstanden werden.

Zufriedenheit ist das wohl am häufigsten verwendete Kriterium für die Bewertung von IS- bzw. WBIS-Erfolg. Der Grund dafür ist, dass Nutzerzufriedenheit relativ einfach zu erfassen ist und als ein guter Indikator für Erfolg angesehen wird. Einen Überblick zu vorhandenen Ansätzen geben DeLone u. McLean (1992); Zviran u. Erlich (2003).

Nutzerzufriedenheit kann in Form eines Gesamturteils wie z. B. „wie schätzen Sie Ihre Zufriedenheit mit [...] ein“ (Rai u. a. 2002, S. 58) erfasst oder mittels verschiedener Kriterien operationalisiert werden. Die Operationalisierung durch das Instrument von Bailey u. Pearson (1983) wird als einer der wichtigsten Beiträge zur Erfassung von IS-Nutzerzufriedenheit im organisationalen Kontext angesehen (Zviran u. Erlich 2003, S. 9). Der Fragebogen besteht aus 39 Fragen zu Management, IT-Abteilung, System und Nutzer. Verschiedene IS- und WBIS-spezifische Ansätze basieren auf diesem Instrument (z. B. Doll u. Torkzadeh 1988; Ives u. a. 1983; McKinney u. a. 2002). Doll u. Torkzadeh (1988) verlagern mit ihrer Adaptation des Instrumentes erstmals den Fokus von der Betrachtung organisationaler IS-Nutzung zu individueller IS-Nutzung („*end-user computing satisfaction*“). Da der Gebrauch von WBIS oft in einem individuellen Kontext erfolgt, fußen viele Ansätze zur Erfassung von WBIS-Zufriedenheit auf dem Instrument von Doll u. Torkzadeh (z. B. Wang u. Tang 2003; Wang u. a. 2001; Xiao u. Dasgupta 2002).

Auch andere Ansätze, die nicht explizit auf Bailey u. Pearson oder Doll u. Torkzadeh basieren, verwenden dennoch gleiche oder ähnliche Kriterien für die Operationalisierung von Zufriedenheit (z. B. Sampson u. Manouselis 2005).

Hauptkritikpunkt an den Operationalisierungsansätzen ist, dass sich die verwendeten Kriterien mit denen für Informations-, System- und Servicequalität überlappen. Daraus ergibt sich die Frage, ob Zufriedenheit überhaupt als eine eigenständige Kategorie angesehen werden kann oder ob es nicht eher ein übergreifendes Gesamtkonstrukt ist, welches sich aus den anderen IS-Qualitätskategorien zusammensetzt (siehe Diskussion in Sedera u. Tan 2005).

Systemnutzung

Neben der Zufriedenheit ist die Systemnutzung eine weitere, oft verwendete IS-Erfolgsbewertungskategorie. Diese wird als eine der objektivsten und am leichtesten erfassbaren angesehen (DeLone u. McLean 1992, 2003).

In der Literatur werden unter anderem Kriterien wie Dauer, Häufigkeit und Umfang der Nutzung verwendet (siehe DeLone u. McLean 1992, 2003). Dass diese nicht immer als Indikatoren für Erfolg geeignet sind, zeigt sich insbesondere bei WBIS. Wenn eine bestimmte Seite einer Website über einen längeren Zeitraum vom Nutzer „betrachtet“ wird, kann das verschiedenen Gründe haben (Sullivan 1997): Der Nutzer ist sehr am Inhalt dieser Seite interessiert; der Nutzer ist für einen gewissen Zeitraum mit etwas anderem beschäftigt; der Nutzer kann auf dieser Seite nicht finden, was er sucht. Das gleiche Problem gilt z. B. für den Umfang der Nutzung. Werden während eines Besuchs auf einer Website besonders viele Seiten anklickt, kann das auf einen besonders interessierten Besucher oder aber er auf einen die gewünschte Information nicht findenden Besucher hinweisen (Sullivan 1997).

Aus der Nutzung ableitbare Performance-Kennzahlen, wie beispielsweise die Konversions² oder die Retentionseffizienz³ einer Website (Berthon u. a. 1996; Cutler u. Sterne 2000) oder die Rate bestimmter komplexer Nutzungsmuster (Berendt u. Spiliopoulou 2000; Spiliopoulou u. Berendt 2001), eignen sich eher als Indikatoren für eine Erfolgsbewertung.

Die Nutzung in die IS-Erfolgsbewertung einzubeziehen, basiert auf der grundlegenden Annahme, dass ein System, das genutzt wird, auch erfolgreich ist (siehe z. B. Davis 1986, 1989). Hier muss man aber zwischen freiwilliger und obligatorischer Nutzung unterscheiden. Bei letzterem kann man nicht immer von obiger Annahme ausgehen.

Systemauswirkungen

Die erfassung und Auswertung der Systemauswirkungen ist der wohl direktteste Ansatz zur IS-Erfolgsbewertung, d. h. es wird letztendlich betrachtet, ob die gesetzten Ziele erreicht wurden. Im Vergleich zu den anderen Bewertungskategorien sind Auswirkungen in der Regel aber auch am schwersten zu erfassen und/oder zu quantifizieren. Mittels welcher Kriterien diese operationalisiert werden können und ob diese überhaupt messbar sind, ist nicht eindeutig festgelegt.

Studien zeigen, dass Auswirkungen sich unter anderem aus der Nutzung des Systems sowie aus der durch die Nutzung resultierenden Nutzermeinung zu Eigenschaften des Systems ergeben (z. B. DeLone u. McLean 1992; Fogg u. a. 2002). Es gibt hierbei mindestens zwei relevante Ebenen (Abschnitt 3.2.5), auf denen die Auswirkungen betrachtet werden können: Individuelle und organisationale Auswirkungen (DeLone u. McLean 1992). Diese können prinzipiell in individuelle und organisationale Kosten und Nutzen aufgegliedert werden (Delić u. Lenz 2006). Abhängig vom Ziel der Evaluation können auch weitere Ebenen wie z. B. Gruppen oder Gesellschaftsschichten in die Untersuchung einbezogen werden (DeLone u. McLean 2003).

²Die Konversionseffizienz ist die Rate der Nutzer, die von Besuchern zu zahlenden Kunden konvertierten.

³Die Retentionseffizienz ist die Rate der wiederkehrenden Kunden.

Individuelle Auswirkungen Der individuelle Nutzen kann unter anderem erfasst werden als verbesserte Entscheidungsfähigkeit, erweitertes bzw. verbessertes Wissen, verbesserte individuelle Produktivität, verbesserte individuelle Effizienz (für IS siehe z.B. DeLone u. McLean 1992; Etezadi-Amoli u. Farhoomand 1996; Igbaria u. Tan 1997; für WBIS siehe z.B. D'Ambra u. Rice 2001).

Vertrauen in ein System, insbesondere beim individuellen Gebrauch von WBIS, entscheidet über Art und Umfang der (zukünftigen) Nutzung des Systems und damit über dessen Erfolg (z.B. Hoffman u.a. 1999; Kim 2000; Torkzadeh u. Dhillon 2002; Warrington u.a. 2000). Vertrauen lässt sich definieren als

the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective to the ability to monitor or control that other party. (Mayer u.a. 1995, S. 712)

Vertrauensbildung wird analog zu den anderen Nutzen- und Kostenfaktoren durch die Eigenschaften des IS sowie durch Erfahrungen des Nutzers mit dem System beeinflusst (z.B. Balasubramanian u.a. 2003; Corritore u.a. 2005; Flavián u.a. 2006; Gefen 2002a; Kim 2000; McKnight u.a. 2002). Die Ausbildung von Vertrauen gegenüber einem System auf Basis der Nutzung und empfundener Qualität kann daher als eine weitere Komponente individuellen Nutzens interpretiert werden.

Die empfundene Unsicherheit im Umgang mit dem System, insbesondere bei öffentlich zugänglichen WBIS (siehe Abschnitt 3.2.8), ist als Risikoempfinden des Nutzers erfassbar (Gefen 2002a; McKnight u.a. 2002). Risiko kann in diesem Zusammenhang definiert werden als

the trustor's belief about likelihoods of gains or losses outside of considerations that involve the relationship with the particular trustee. (Mayer u.a. 1995, S. 726)

Je größer also das Vertrauen eines Individuums in die gegenüberliegende Partei im Vergleich zum empfundenen Risiko ist, desto größer sollte dessen Bereitschaft sein, ein entsprechendes risikobehaftetes Verhältnis einzugehen (Mayer u.a. 1995). Im Rahmen der IS-Nutzung lässt sich diese Aussage so interpretieren: Die Bereitschaft ein IS zu nutzen, hängt vom Verhältnis zwischen Vertrauen in das System bzw. die das System betreuende Partei und dem empfundenen Risiko der Nutzung des Systems ab. Das empfundene Risiko wird damit zu einem individuellen Kostenfaktor.

Positive Erfahrungen mit einem System steigern das entgegengebrachte Vertrauen, senken das empfundene Risiko und stärken langfristig die Bindung zu dem System bzw. die Loyalität des Nutzers (Flavián u.a. 2006; Gefen 2002a; Lee u.a. 2000b). Loyalität kann definiert werden als

a deeply held commitment to rebuy or repatronize a preferred product/service consistently in the future, thereby causing repetitive same-brand or same brand-set purchasing, despite situational influences and marketing efforts having the potential to cause switching behavior. (Oliver 1999, S. 34)

Loyalität wird als einer der wichtigsten Erfolgsfaktoren im eCommerce gesehen (Reicheld u. Shefter 2000). Vertrauen und Loyalität steigern das individuelle Sicherheitsgefühl des Nutzers und festigen die Beziehung zum Provider des IS bzw. WBIS. Zusätzlich kann der Nutzer in so einer Beziehung von speziellen Angeboten und Dienstleistungen profitieren, die für Außenstehende nicht erhältlich sind. Loyalität lässt sich damit als individueller Nutzenfaktor interpretieren.

Von der Stärke der Loyalität hängt ab, wie groß die Bereitschaft zu Aufwendungen für die Suche nach einem neuen System, das Eingehen einer neuen risikobehafteten Beziehung und zur Einarbeitung in das System ist. Solche Systemwechselkosten (Gefen 2002a) können daher als ein weiterer individueller Kostenfaktor interpretiert werden.

Organisationale Auswirkungen Vertrauen und Loyalität führen zu bevorzugter und regelmäßiger Nutzung des Systems (z.B. Onlineshop), woraus sich wiederum positive organisationale Effekte für den Provider ergeben können, wie z.B. gestiegener Umsatz und damit verbesserte Wettbewerbsfähigkeit (Warrington u.a. 2000).

Der Nutzen eines Systems kann unter anderem auf Basis finanzieller Kennzahlen, wie z.B. Profit, Rentabilität und Kostensenkungspotential, wie auch auf Basis nicht-monetärer Performance-Kennzahlen, wie z.B. Produktivität, Innovationspotential, Wettbewerbsfähigkeit oder Prestige, erfasst werden (siehe DeLone u. McLean 1992, 2003; Mirani u. Lederer 1998).

Analog dazu können Kosten ebenfalls mittels finanzieller oder nicht-monetärer Kennzahlen ermittelt werden. Die Gesamtbetriebskosten (engl.: *Total Cost of Ownership* (TCO)) lassen sich in die Hauptkategorien „Aquisitionskosten“ und „Betriebskosten“ unterteilen (David u.a. 2002; van Maanen u. Berghout 2002). Aquisitionskosten sind alle mit der Anschaffung und in Betriebnahme des Systems verbundene Kosten (z.B. Kosten für Hard- und Software). Betriebskosten sind alle zum laufenden Betrieb des Systems notwendigen Ausgaben (z.B. für Personal oder Strom). Zur Erfassung letzterer können auch nicht-monetäre Kennzahlen eingesetzt werden, wie z.B. Systemausfallzeiten (David u.a. 2002) oder die Rate von ehemaligen Nutzern, die das System nicht mehr benutzen (Cutler u. Sterne 2000).

Modelle relevanter Variablen von IS-Erfolg

Die vorangegangenen Abschnitte enthalten eine einführende Übersicht zu verwendeten Bewertungskriterien für IS bzw. WBIS. Wie oben zu ersehen, ist die Anzahl und Verschiedenheit entsprechender Kriterien groß (siehe DeLone u. McLean 1992; Grover u. a. 1996; Seddon u. a. 1999). Dementsprechend ist es schwierig oder gar unmöglich, im Rahmen einer Bewertung alle in Frage kommenden Kriterien zu identifizieren und einzubeziehen (siehe Erläuterung in Abschnitt 3.1). Es muss somit eine Auswahl der in Bezug auf das konkret betrachtete IS als relevant erachteten vorgenommen werden. In der Literatur finden sich diesbezüglich verschiedene Ansätze, die in entsprechenden Modellen die Zusammenhänge zwischen als relevant erachteten Variablen von IS-Erfolg postulieren. Drei im IS-Forschungsfeld verbreitete Modelle (Tabelle 3.1) sind das *Technology Acceptance Model* (TAM) von Davis (1986, 1989), das *IS Success Model* (D/M92) von DeLone u. McLean (1992) und das *Task-Technology Fit* (TTF)-Modell von Goodhue u. Thompson (1995). Diese Modelle stellen drei verbreitete Ansichten zur „Abhängigen Variable“ bzw. zu Hauptkriterien von IS-Erfolg dar:

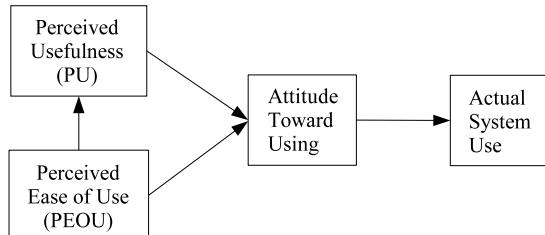
- Akzeptanz bzw. Nutzung des Systems (Davis 1986, 1989),
- individuelle, organisationale und andere Auswirkungen des Systems (DeLone u. McLean 1992, 2003),
- Passgenauigkeit (engl.: *fit*) zwischen auszuführender Aufgabe und Technologie des Systems (Goodhue u. Thompson 1995).

Tabelle 3.1: Modelle relevanter Variablen von IS-Erfolg

Modell	Unabhängige Variablen	Abhängige Variablen	Ebenen der Analyse
TAM	<i>Perceived Ease of Use,</i> <i>Perceived Usefulness</i>	<i>Behavioral Intention to Use,</i> <i>Actual System Use</i>	Individuum
D/M92	<i>Information Quality,</i> <i>System Quality</i>	<i>Use,</i> <i>User Satisfaction,</i> <i>Individual Impact,</i> <i>Organizational Impact</i>	System, Individuum, Organisation
TTF	<i>Task Characteristics,</i> <i>System Characteristics</i>	<i>Task-Technology Fit,</i> <i>Utilization,</i> <i>Performance Impacts</i>	Individuum

Technology Acceptance Model (TAM) Das TAM ist eine Adaption der *Theory of Reasoned Action* (TRA) von Fishbein u. Ajzen (1975) auf IS. TRA postuliert, dass individuelles Verhalten durch Verhaltensabsichten gesteuert wird, welche wiederum von individueller Einstellung und subjektiven Normen beeinflusst werden. Das TAM besagt entsprechend, dass die Absicht eines Individuums zur

(zukünftigen) Nutzung eines IS abhängig von empfundener Brauchbarkeit (engl.: *Perceived Usefulness* (PU)) und empfundenen Bedienkomfort (engl.: *Perceived Ease of Use* (PEOU)) ist. Die Absicht wiederum beeinflusst dann die tatsächliche Nutzung (Abbildung 3.5).



Quelle: Davis (1986)

Abbildung 3.5: *Technology Acceptance Model*

Stärke des TAM ist sein effizienter Ansatz zur Erfassung individueller Faktoren der Akzeptanz bzw. des Erfolgs von IS. Die rege TAM-basierte Folgeforschung ist ein Indikator für das dementsprechend große Interesse an diesem Modell in der IS-Forschung.

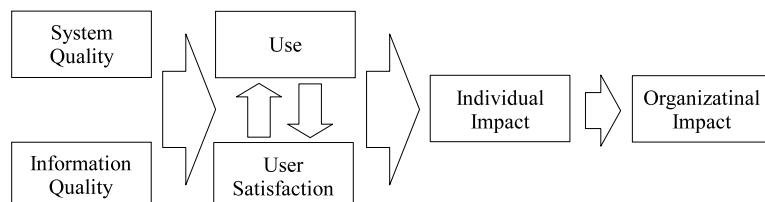
Schwäche des Ansatzes ist, dass nur Technologie und Design eines Systems betrachtet werden, ohne andere Faktoren zu berücksichtigen. So basieren TRA und TAM auf der Annahme, dass auf eine Absicht zur Handlung immer auch eine Handlung folgen kann. In der Realität dagegen treten Barrieren auf, wie z. B. beschränkte individuelle Fähigkeiten, zeitliche Beschränkungen, Kapazitätsbeschränkungen, organisationale Beschränkungen und beschränkende Umwelteinflüsse. Daneben erfolgt die Nutzung eines Systems nicht immer in freiwilligem Rahmen, wodurch Nutzung als Erfolgskriterium aufgeweicht wird.

Auf dem TAM basieren zahlreiche Folgestudien, die das Modell validieren, anwenden und erweitern (siehe Lee u. a. 2003; Venkatesh u. a. 2003). Die Modifikationen beinhalten IS- und Kontextspezifische Erweiterung um externe und moderierende Variablen der PU- und PEOU-Komponenten (siehe Lee u. a. 2003). Basierend auf den in Folgestudien identifizierten Beschränkungen des Modells, konzipieren Venkatesh (2000) und Venkatesh u. Davis (2000) das TAM II. Dieses stellt eine Erweiterung des alten Models um relevante externe und moderierende Variablen dar.

IS Success Model (D/M92) Das D/M92 von DeLone u. McLean (1992) basiert auf der *Communications Theory* von Shannon u. Weaver (1949) und Masons (1978) Adaption der Theorie auf IS. Shannon u. Weaver postulieren, dass „Information“ in einem Kommunikationssystem auf verschiedenen Ebenen gemessen werden kann: Auf einer technischen, einer semantischen und auf einer Ebene der Effektivität. Nach Mason durchläuft „Information“ mehrere verschiedene Stufen von dessen Produktion über dessen Gebrauch bis hin zu dessen Einfluss auf individuelle

und/oder organisationale Performance. Daraus folgert er, dass für jede Ebene, auf der „Information“ transportiert wird, eine separate Erfassung bzw. Analyse für die Erfolgsbewertung eines IS nötig sei.

Auf Grundlage dieser zwei Studien und der Auswertung verwendeter IS-Bewertungskriterien in über 180 Arbeiten aus den Jahren 1981–87 identifizieren DeLone u. McLean (1992) sechs interdependente Basiskategorien von IS-Erfolg (Abbildung 3.6): (1) Systemqualität (engl.: *system quality*), (2) Informationsqualität (engl.: *information quality*), (3) (System-)Gebrauch (engl.: *use*), (4) Nutzerzufriedenheit (engl.: *user satisfaction*), (5) Individuelle Auswirkungen (engl.: *individual impact*) und (6) Organisationale Auswirkungen (engl.: *organizational impact*). Um die Relevanz dieser Kategorien zu betonen, bezeichnen sie diese als „Dimensionen“ von IS-Erfolg.



Quelle: DeLone u. McLean (1992)

Abbildung 3.6: *IS Success Model*

Zu den Stärken des D/M92 zählt, dass es IS-Erfolg als ein multikategorisches bzw. „multidimensionales“ Konstrukt versteht. Auf Basis des Modells ist der Einbezug verschiedener Erfolgskategorien auf verschiedenen Ebenen (Individuum, Organisation, System) im Rahmen einer Analyse möglich. Dieses kann zur Gewinnung sehr umfangreicher und detaillierter Analyseergebnisse hilfreich sein. Außerdem besitzt das Modell damit eine gewisse Flexibilität, wodurch dessen Einsatz in vielen verschiedenen Realweltkontexten möglich ist.

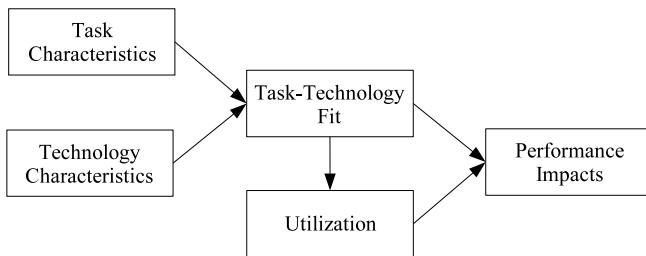
Als eine der Hauptschwächen wird die Vermengung von kausalen und prozessualen Abhängigkeiten in dem Modell gesehen, d. h. die dargestellten Beziehungen zwischen den Kategorien können sowohl als Abhängigkeiten interpretiert werden als auch als Prozess. Ein Kausalmodell basiert auf der Annahme, dass für eine betrachtete Datenmenge, ceteris paribus, die Varianz einer unabhängigen (Modell-) Variablen notwendig und ausreichend ist, um die Varianz einer abhängigen Variablen zu beeinflussen. Im Gegensatz dazu zeigt ein Prozessmodell, wie eine Sequenz von Ereignissen ein bestimmtes Endresultat bewirkt. Ein Ereignis in einer Sequenz ist, für sich genommen, notwendig aber nicht hinreichend, um das Endresultat zu bewirken. Daneben ist die Bedeutung der Kategorie *use* verschieden auslegbar: (1) als aktueller Indikator für Erfolg, (2) als Indikator für zukünftigen Gebrauch des Systems oder (3) als Schritt in einem Prozess, der zu individuellen und organisationalen Auswirkungen führt. Diese Unklarheiten schwächen die Aussagekraft des Modells (Seddon 1997).

Basierend auf zahlreichen Folgestudien, die das Modell validieren, anwenden und erweitern (siehe DeLone u. McLean 2003), entwickeln DeLone u. McLean (2003) eine respezifizierte Version (*IS Success Model 2003* (D/M03)). Diese ist um die Kategorie „Servicequalität“ (engl.: *service quality*) erweitert, um der zunehmenden Bedeutung von Servicequalität insbesondere im Rahmen der individuellen Nutzung eines Systems Rechnung zu tragen. Daneben sind die Kategorien „individuelle Auswirkungen“ und „organisationale Auswirkungen“ des ursprünglichen Modells durch eine Kategorie „Nettonutzen“ (engl.: *net benefits*) ersetzt, da IS auch außerhalb des individuellen und organisationalen Bereichs erfolgsrelevante Auswirkungen haben können.

Task-Technology Fit (TTF) Das TTF-Modell ist Teil des *Technology-to-Performance Chain* (TPC)-Modells von Goodhue u. Thompson (1995). Die Kernannahme des TPC-Modells ist, dass Technologie nur dann positive Auswirkungen auf die individuelle Performance eines Nutzers haben kann, wenn diese zum einen vom Individuum genutzt wird und zum anderen eine gute Entsprechung (engl.: *fit*) zwischen zu bewältigender Aufgabe, individuellen Fähigkeiten und den Eigenschaften der eingesetzten Technologie vorliegt.

Das TPC-Modell kombiniert zwei IS-Forschungsansätze: (1) Nutzereinstellung und -absichten als Einflussfaktoren auf die tatsächliche Nutzung eines Systems (Davis 1986, 1989) und (2) die Entsprechung zwischen Aufgabe und verwendeter Technologie (engl.: *task-technology fit*) als Einflussfaktor auf die erzielbare Performance eines Individuums (Goodhue 1988).

Goodhue u. Thompson validieren nicht das komplette TPC-Modell, sondern nur einen Teil (Abbildung 3.7). Dieser besteht aus den Kategorien (1) Charakteristiken der Aufgabe (engl.: *task characteristics*), (2) Charakteristiken der Technologie (engl.: *technology characteristics*), (3) Entsprechung zwischen Technologie und Aufgabe (engl.: *task-technology fit*), (4) Gebrauch der Technologie (engl.: *utilization*) und (5) Performanceauswirkungen (engl.: *performance impacts*). In der Literatur wird dieses Teilmodell oft als „Task-Technology Fit-Modell“ bezeichnet (z.B. Dishaw u. Strong 1999).



Quelle: Goodhue u. Thompson (1995)

Abbildung 3.7: Task-Technology Fit-Modell

Stärke des Modells ist seine erweiterte Betrachtung von Einflussfaktoren auf

die individuelle Performance: Erfolgreicher Einsatz von Technologie benötige neben dessen Nutzung auch eine gewisse entsprechung mit den charakteristiken der durchzuführenden Aufgabe.

Folgearbeiten beinhalten neben der Anwendung (z.B. D'Ambra u. Rice 2001) und Erweiterung des Modells (z.B. Dishaw u. Strong 1999; Goodhue u.a. 2000; Venkatesh u.a. 2003) auch eine empirische Validierung verschiedener Teilbereiche. So können beispielsweise Staples u. Seddon (2004) den postulierten Einfluss individueller Charakteristiken des Users auf die System-Performance bestätigen.

3.3 Verbreitete Ansätze zur Evaluation von Websites

Bezug nehmend auf das in Abschnitt 3.2 vorgestellte Rahmenmodell zur Evaluation von IS – der IS-Erfolgspyramide – werden hier gängige Bewertungsansätze für Websites vorgestellt. In der Literatur sind drei Ansätze besonders häufig vertreten: (1) Die Analyse der Nutzermeinungen, (2) die Analyse des Nutzerverhaltens und (3) die expertenbasierte Analyse der Websitequalität.

Die Analyse der Nutzermeinungen erfolgt in der Regel auf der individuellen Ebene des Nutzers per Fragebogen. Die Analyse des Nutzerverhaltens wird meist auf der individuellen Ebene per Fragebogen oder auf Systemebene per Logdatenanalyse durchgeführt. Die expertenbasierte Analyse wird entweder auf der individuellen Ebene per Fragebogen, manueller Beobachtung oder Rollenspiel durchgeführt, oder auf der Systemebene mittels automatisierter Beobachtung (Tabelle 3.2).

Tabelle 3.2: Verbreitete Ansätze zur Evaluation von Websites

Bewertungskriterium	Analyseebene	Methode der Datenerfassung und -analyse
Nutzermeinung	Individuum	Fragebogen
Nutzerverhalten	Individuum	Fragebogen
Nutzerverhalten	System	Beobachtung (von Logdaten)
Expertemeinung	Individuum	Beobachtung (manuell), Fragebogen, Simulation (Rollenspiel)
Expertemeinung	System	Beobachtung (automatisch)

3.3.1 Analyse der Nutzermeinungen

Die wohl am häufigsten genutzte Methode zur Analyse von Nutzermeinungen ist die Umfrage per Fragebogen. Bestehende Fragebogenansätze lassen sich entsprechend der im vorangegangenen Abschnitt vorgestellten Modelle relevanter Variablen von IS-Erfolg unter anderem in drei Kategorien untergliedern (Abbildung 3.8): (1) TAM-basierte, (2) TTF-basierte sowie (3) D/M92- bzw. D/M03-basierte. Zu letzterer Kategorie können auch Fragebogenansätze gezählt werden, die Kriterien verwenden, die einer oder mehreren der in den Modellen von DeLone u. McLean definierten IS-Erfolgskategorien zugeordnet werden können.

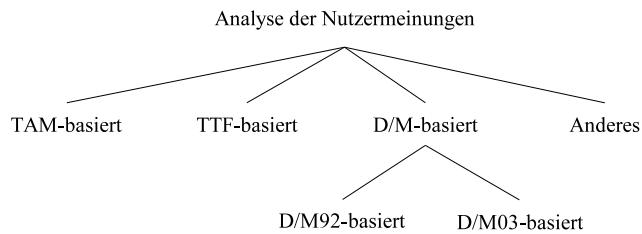


Abbildung 3.8: Ansätze zur Analyse von Nutzermeinungen

Bestehende Ansätze sind entweder explizit als Instrumente zur Analyse von Websites konzipiert (z.B. Aladwani u. Palvia 2002) oder aber dienen in erster Linie zur Validierung von in Modellen postulierten Zusammenhängen zwischen Variablen für spezifische Typen von Websites (z.B. van der Heijden 2003).

TAM-basierte Ansätze wenden das Modell im Kontext des Web an und erweitern es um Konstrukte, wie z.B. den Unterhaltungsfaktor (Chung u. Tan 2004; Hwang u. Yi 2002; Koufaris 2002; Moon u. Kim 2001), die Selbstwirksamkeit (engl.: *self efficacy*) des Nutzers (Hwang u. Yi 2002), die empfundene Kontrolle über das System (Koufaris 2002), die visuelle Attraktivität (van der Heijden 2003), den Typ der Aufgabenstellung (Speier u. Poston 2001), empfundenes Vertrauen und Risiko (Gefen 2002b; Gefen u. a. 2003; Pavlou 2001; Shih 2004), den technologischen Wandel im Laufe der Zeit (Wang u. Butler 2003), Informationsbeschaffung und Kauf online (Gefen u. Straub 2000), die Servicequalität (Lin u. Wu 2002), individuelle Einstellungen bzw. Neigungen der User (Shih 2004) sowie deren kultureller Hintergrund (Kralisch 2005).

Die in TTF-basierten Ansätzen verwendeten TTF-Faktoren im Kontext des Web sind unter anderem: Unterhaltungsfaktor, Weberfahrung, Auffindbarkeit von Information, Shoppingkosten, Kontrolle über das System, soziale Einflüsse (z.B. D'Amra u. Rice 2001). Erweitert bzw. modifiziert wird das Modell mit Konstrukten, wie z.B. der Unsicherheit bzw. dem empfundenen Risiko (D'Amra u. Wilson 2004), nationalen und interorganisationalen kulturellen Einflüssen (Vatanasakdakul u. D'Amra 2007), den sozialen Bedürfnissen des Users (Kwai Fun IP u. Wagner 2008), den individuellen Eigenschaften des Users (Lee u. a. 2007).

Auf dem D/M92 bzw. D/M03 basierende Ansätze betrachten meist nur eine Auswahl der von DeLone u. McLean (1992, 2003) identifizierten IS-Erfolgskategorien (siehe Tabelle 3.3). Daneben gibt es viele Ansätze, die nicht explizit auf einem der D/M-Modelle basieren, aber Kriterien verwenden, die einer oder mehreren der von DeLone u. McLean identifizierten IS-Erfolgskategorien zugeordnet werden können (siehe Tabelle 3.4).

Prinzipiell können die meisten IS-Bewertungskriterien (siehe Abschnitt 3.2.8) im Rahmen einer Befragung verwendet werden. Zwischen den auf verschiedenen Bewertungsmodellen basierenden Ansätzen gibt es hierbei Überlappungen. So entspricht beispielsweise die Kategorie *use* des TAM der Kategorie *use* der D/M-

Tabelle 3.3: Ansätze die auf dem D/M92 bzw. D/M03 basieren

Autor(en)	IQ/CQ	SQ	SRVQ	ZF	NU	A
Landrum u. Prybutok (2004)	✓	✓	✓	✓◊		✓
McKinney u. a. (2002)	✓	✓		✓◊		
Molla u. Licker (2001)	✓	✓	✓	✓◊	✓	✓
Negash u. a. (2003)	✓	✓	✓	✓◊		
Webb u. Webb (2002)	✓		✓			

✓ vorhanden
 IQ/CQ Informations-/Contentqualität
 SRVQ Servicequalität
 NU Nutzung

◊ abhängige Variable des Modells
 SQ Systemqualität
 ZF Zufriedenheit
 A Anderes

Modelle. Die Kategorie *performance* des TTF-Modells entspricht den *impact* Kategorien des D/M92. Daneben werden auch gleiche oder ähnliche Kriterien verwendet. So z. B. bei Kriterien, die zur Erfassung von Bedienkomfort (*Perceived Ease of Use*) im TAM, Systemqualität (*system quality*) in den D/M-Modellen und Eigenschaften der Technologie (*technology characteristics*) im TTF-Modell verwendet werden.

Neben vielen Übereinstimmungen werden in diesen drei Modellen durch ihre unterschiedlichen Schwerpunkte aber auch sehr spezifische, eigenständige Kriterien verwendet. Aus diesem Grund kombinieren einige Ansätze die Modelle oder deren Komponenten, um so ein erweitertes Analyseinstrument zu bekommen. So z. B. TAM und Servicequalität (Devaraj u. a. 2002), TAM und TTF (Dishaw u. Strong 1999; Klopping u. McKinney 2004; Pagani 2006) oder TAM und Informations-, System- und Servicequalität (Cao u. a. 2004).

3.3.2 Analyse des Nutzerverhaltens

Zwei verbreitete Ansätze zur Analyse des Nutzerverhaltens sind die Befragung der Nutzer und die Beobachtung von Nutzerverhalten durch Auswertung von Webserver-Logdaten. Mittels Befragung können nur sehr grundlegende Informationen, wie z. B. Nutzungshäufigkeit und durchschnittliche Nutzungsdauer eines Systems, erfasst werden (z. B. Moon u. Kim 2001). Es ist so gut wie ausgeschlossen, dass ein Nutzer darüber hinaus für einen längeren Zeitraum in der Vergangenheit genaue Angaben zu seinen Klickpfaden auf einer Website machen kann. Außerdem können Befragungsergebnisse von den wahren Verhaltensmustern abweichen, da Nutzer ihr eigenes Verhalten nicht immer genau einschätzen können (Deane u. a. 1998). Die Logdatenauswertung dagegen erlaubt eine präzisere Erfassung des Nutzerverhaltens.

Bestehende Ansätze zur Auswertung von Logdaten können in zwei große Gruppen gegliedert werden: (1) Verfahren die auf *Data Mining* basieren und in diesem Zusammenhang als *Web Usage Mining*-Verfahren bezeichnet werden, (2) und Ver-

Tabelle 3.4: Ansätze die von DeLone u. McLean identifizierte IS-Erfolgskategorien bzw. -kriterien verwenden

Autor(en)	IQ/ CQ	SQ	SRQ	ZF	VE	RI	LO	KO	EI	A
Aladwani u. Palvia (2002)	✓	✓								
Appan u. Mellarkod (2003)	✓	✓			✓◊					
Balasubramanian u. a. (2003)			✓	✓◊	✓			✓	✓	✓
Corritore u. a. (2005)		✓			✓◊	✓				✓
Gefen (2002a)			✓		✓	✓	✓	✓◊		
Grigoroudis u. a. (2008)	✓	✓								
McKnight u. a. (2002)			✓		✓				✓	✓
Muylle u. a. (2004)	✓	✓		✓◊						
Ridings u. a. (2002)					✓					✓
van Riel u. a. (2001)			✓	✓			✓			
Sampson u. Manouselis (2005)	✓	✓								
Wang u. Tang (2003)				✓						
Wang u. a. (2001)	✓	✓	✓		✓◊	✓				
Woodroof u. Burg (2003)					✓					✓
Xiao u. Dasgupta (2002)	✓	✓			✓◊					
Yoo u. Donthu (2001)	✓				✓					

✓	vorhanden	◊	abhängige Variable des Modells
IQ/CQ	Informations-/Contentqualität	SQ	Systemqualität
SRQ	Servicequalität	ZF	Zufriedenheit
VE	Vertrauen	RI	Risiko
LO	Loyalität	KO	Kosten
EI	Einstellung	A	Anderes

fahren zur (simplen) statistischen Auswertung, wie z.B. die Zugriffshäufigkeiten auf eine Seite.

Die im Folgenden verwendeten Begriffe im Kontext des World Wide Web entsprechen den Definitionen des W3C-Konsortiums (siehe Beschreibungen in Cooley 2003; Lavoie u. Nielsen 1999) und sind in Tabelle 3.5 kurz erläutert.

Web Usage Mining

Web Usage Mining ist ein Teilgebiet des *Web Mining* (Abbildung 3.9), welches die Anwendung von Data Mining-Methoden zur Extraktion von Informationen aus und im Kontext des WWW zum Gegenstand hat (Kosala u. Blockeel 2000). Data Mining oder auch *Knowledge Discovery in Databases* ist die nichttriviale Extraktion impliziter, vorher unbekannter und potenziell nützlicher Informationen aus Datenbeständen (Frawley u. a. 1992). Gegenstand des Web Usage Mining ist die Analyse der Nutzungsmuster von Websitebesuchern. Die Ergebnisse können

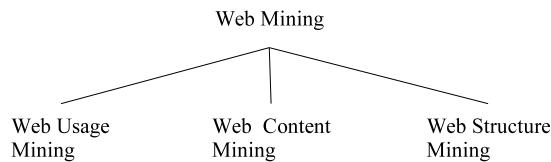


Abbildung 3.9: Taxonomie des Web Mining

z. B. dazu verwendet werden, Inhalte zu personalisieren, das zugrunde liegende System einer Website und/oder die vom Nutzer sichtbare Oberfläche und Struktur zu verbessern, *Business Intelligence* (BI)-Informationen zu extrahieren oder Web-Nutzung zu charakterisieren (Srivastava u. a. 2000).

Das Nutzerverhalten wird aber erst analysier- und interpretierbar, wenn es im Kontext von Inhalten und Strukturen der besuchten Websites und abgerufenen Seiten betrachtet wird (Cooley 2003). Die Auswertung von Inhalten und Strukturen ist Gegenstand zweier weiterer Teilgebiete des Web Mining: *Web Structure Mining* und *Web Content Mining* (Kosala u. Blockeel 2000; Srivastava u. a. 2000).

Gegenstand des Web Content Mining ist die Analyse von Seiteninhalten. Diese können aus einfachem Text, Metadaten, Audiodaten, Videodaten, Hyperlinks u. ä. bestehen. Mittels Web Content Mining können Seiten beispielsweise auf Basis ihrer Inhalte gruppiert bzw. bestimmten Themengebieten zugeordnet werden. So ein Vorgehen ist z. B. für Suchvorgänge interessant. Angewandte Methoden sind unter anderem die Clusterung (Dhyani u. a. 2002; Zamir u. a. 1997) oder die Klassifikation (Pirolli u. a. 1996). Andere Ansätze gehen über eine Verhaltensanalyse der Website-Besucher, basierend auf der Annahme, dass ein Nutzer während seines Besuchs von der Thematik ähnliche Seiten anklickt. Seiten, die von vielen Besuchern zusammen angeklickt wurden, werden dann derselben Gruppe zugeordnet (Dhyani u. a. 2002).

Gegenstand des Web Structure Mining ist zum einen die Analyse von Datenstrukturen innerhalb einer Seite (engl.: *intra page structure*) und zum anderen die Analyse von Verlinkungsstrukturen zwischen Seiten innerhalb einer Website oder zwischen verschiedenen Websites (engl.: *inter page structure*), siehe z. B. Gillenson u. a. 2000; Mukherjea u. a. 1995; Pirolli u. a. 1996.

Der Web Usage Mining Prozess lässt sich analog zum Data Mining Prozess in vier Phasen aufteilen (Kosala u. Blockeel 2000, S. 2): (1) Datenbeschaffung, (2) Datenselektion und -aufbereitung, (3) Musterextraktion und (4) Musterinterpretation (Abbildung 3.10).

Datenbeschaffung Benötigte Datenquellen werden identifiziert und die Daten extrahiert. Web-Nutzungsdaten können auf drei Ebenen erfasst werden: Auf Webserver-, Proxyserver- und Client-Ebene (Srivastava u. a. 2000).

Auf Webserver-Ebene (engl.: *server-side*) werden Nutzungsdaten in Logdateien (W3C 1995) gespeichert. In Abhängigkeit von der Serverkonfiguration kön-

Tabelle 3.5: Web-Begriffe

Begriff	Beschreibung
User (Nutzer)	Ein <i>User</i> oder zu Deutsch „Nutzer“ ist ein einzelnes Individuum, welches mittels Webbrower mit Webserversystemen interagiert.
Web page (Webseite)	Eine <i>Web page</i> (auch <i>Webpage</i>) oder zu Deutsch „Webseite“ bzw. „Seite“ ist ein Verbund aus mehreren Websourcen (z. B. Html-Datei und Grafiken) die bei Abruf dieser Seite zusammen auf dem Webbrower des Users erscheinen sollen und die durch eine einheitliche Seitenadresse gekennzeichnet sind.
Web site (Webpräsenz)	Eine <i>Web site</i> (auch <i>Website</i>) oder zu Deutsch „Webpräsenz“ ist eine Sammlung von miteinander verbundenen Seiten, die unter einer einheitlichen Webadresse erreichbar sind.
Page view	Ein <i>Page view</i> ist die visuelle Wiedergabe einer Seite auf dem Webbrower des Nutzers.
Click-stream	Ein <i>Click-Stream</i> ist eine Sequenz von Page-View-Anfragen.
User-Session	Eine <i>User-Session</i> ist der Click-Stream eines Nutzers über das ganze Web während eines Besuchs.
Server-Session	Eine <i>Server-Session</i> ist der Click-Stream eines Nutzers auf einer Webpräsenz während eines Besuchs.
Episode	Eine <i>Episode</i> ist eine semantisch bedeutungsvolle Teilmenge einer Session.
Download	Ein <i>download</i> beinhaltet das Herunterladen eines sich auf dem Webserver befindlichen Dokumentes (z. B. *.pdf oder *.jpeg) auf den lokalen Rechner des Users.

nen diese verschiedene Informationen, wie z. B. die *Internet Protocol (IP)*-Adresse eines zugreifenden Nutzers, Datum und Uhrzeit der Anfrage sowie die *Uniform Resource Identifier (URI)*-Adresse des angeforderten Webdokumentes, enthalten (Hallam-Baker u. Behlendorf 2002). Zu Erläuterungen dieser Begriffe und weiteren wichtigen Feldern einer Webserver-Logdatei siehe Tabelle 3.6.

Ein Proxyserver (Proxyserver-Ebene, engl.: *proxy-side*) steht im Internet zwischen Nutzer und Webserver und leitet Nutzeranfragen an die entsprechenden Webserver weiter. Die Auswertung von Proxysserver-Logdaten erlaubt somit eine Analyse des Nutzerverhaltens über alle Websites, die der Nutzer während seines Aufenthalts im Web (Session) besucht hat (Srivastava u. a. 2000).

Die Erfassung auf Client-Ebene (engl.: *client-side*) wird durch entsprechende Programme auf dem Rechner des Nutzers ermöglicht. So wird nicht nur sein gesamtes Surfverhalten, sondern auch sein Bedienverhalten der Browsersoftware erfassbar. Es lässt sich z. B. nachvollziehen, wann und wie oft der „Zurück“-Button des Browsers gedrückt wurde und ob dies ein wesentlicher Bestandteil des Nutzungsverhaltens dieses Users ist (Catledge u. Pitkow 1995).

Für den Anbieter einer Website bzw. eines Webportals sind in der Regel nur die

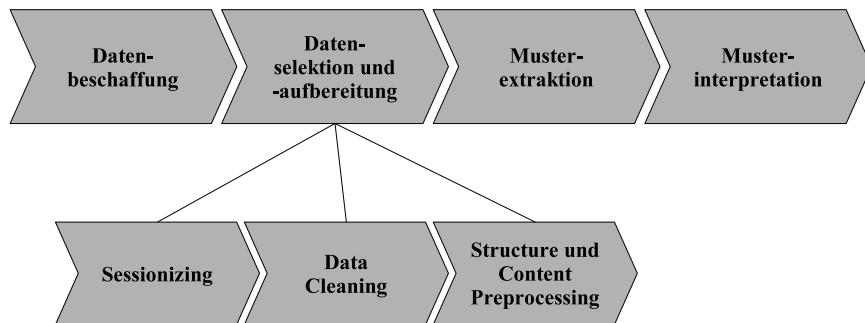


Abbildung 3.10: Web Usage Mining Prozess

eigenen Server-Logdaten als Quelle einer Analyse von Nutzungsmustern zugänglich. Die im Folgenden beschriebenen Schritte beziehen sich daher auf Logdaten dieses Typs.

Im Originalzustand sind Logdaten für Web Usage Mining wenig brauchbar. Die interessierenden bzw. benötigten Inhalte müssen dafür zunächst extrahiert und aufbereitet werden. Das geschieht in der zweiten Phase des Web Usage Mining Prozesses.

Datenselektion und -aufbereitung Die Selektion und Aufbereitung der Daten (engl.: *preprocessing*) besteht im Wesentlichen aus drei Schritten (Cooley u. a. 1999; Facca u. Lanzi 2005): (1) Rekonstruktion von Besuchersitzungen (engl.: *sessionizing*), (2) Bereinigung der Logdaten von irrelevanten Einträgen und Sessions (engl.: *data cleaning*) und (3) Vorverarbeitung von für die Untersuchung relevanten Inhalts- und Strukturinformationen (engl.: *structure and content preprocessing*). Schritt 2 kann abhängig von der Untersuchung auch vor Schritt 1 ausgeführt werden.

Sessionizing Ein Klick auf einen Link in einer Seite resultiert im Download von mehreren Dateien auf den Browser des Nutzers (Page View),⁴ wie z. B. der eigentlichen Html-Datei, eingebetteten Grafik-, Text-, Multimediateilen usw. Jede einzelne dieser angeforderten Dateien erzeugt einen Eintrag im Log des Webservers. In der Regel werden keine zusätzlichen Informationen darüber, welche Einträge zu einem Page View gehören und welche Page Views zu einer Session gehören, festgehalten. Diese Zusammenhänge müssen rekonstruiert werden. Die Qualität der Rekonstruktion hängt von den im Log vorhandenen Informationen und von der angewendeten Session-Rekonstruktionsstrategie bzw. vom serverseitig bereitgestellten Session-Rekonstruktionsmechanismus ab (Srivastava u. a. 2000). Bei entsprechend vorhandenen Informationen ist im Idealfall eine nahezu vollständige Rekonstruktion möglich (Berendt u. a. 2001; Spiliopoulou u. a. 2003). In der Regel ist dieses

⁴Siehe Erläuterung in Tabelle 3.5.

Tabelle 3.6: Wichtige Felder eines Webserver-Logs

Feld	Beispiel	Beschreibung
IP	123.456.789.999	Eindeutige Internetadresse des Rechners, von dem aus der Zugriff auf den Webserver erfolgte.
Time	[10/Oct/2000:13:55:36 -0700]	Datum und Uhrzeit des Zugriffszeitpunktes sowie die Zeitzone aus der der Zugriff stammt.
Method	GET	Zugriffsmethode auf ein Webdokument.
URI	/ordner/a.html	Einheitliche Adresse für Ressourcen auf der Website. Beinhaltet den Standort und Namen des angefragten Webdokumentes auf dem Webserver.
Protocol	HTTP/1.0	Das vom Client verwendete HTTP-Protokoll.
Statuscode	200	Der Statuscode zu einer Anfrage gibt an, ob diese erfolgreich ausgeführt wurde bzw. welcher Fehler auftrat.
Size	2360	Größe des angefragten Webdokumentes in Kilobyte.
Referrer	http://www.website.com/wo_ich_zuletzt_war.html	Website und Seite auf der sich der Nutzer zuletzt befand und von der aus die aktuelle Seite angefragt wurde.
Agent	Mozilla/4.08 [en] (Win98; I ;Nav)	Vom Webbrower des Users gesendete Kennung.

aber aus verschiedenen Gründen nicht der Fall (Cooley u. a. 1999; Srivastava u. a. 2000):

- Ohne Login-Mechanismus lässt sich nicht mit absoluter Sicherheit feststellen, wann die Session eines Users beendet wurde und wann die nächste angefangen hat. Üblicherweise wird eine Maximaldauer von 30 Minuten für eine Session angesetzt (Spiliopoulou u. a. 2003).
- Nutzer die gemeinsam über einen Proxyserver zugreifen, hinterlassen die gleiche IP-Adresse im Log. Ihre Zugriffe können ohne eindeutige Identifizierung, wie z. B. Cookies,⁵ nicht von einander unterschieden werden.
- Browser und Proxyserver können Webdokumente zwischenspeichern. Der Sinn der Zwischenspeicherung liegt darin, häufig angeforderte Dokumente schnell verfügbar zu machen, ohne jedes Mal zeit- und netzlastaufwändig auf

⁵Cookies enthalten einen eindeutigen Code, der bei jedem Zugriff übertragen wird und so einen Nutzer wiedererkennbar macht.

einen Webserver zugreifen zu müssen. Wird ein zwischengespeichertes Dokument angefordert, erfolgt keine Anfrage (engl.: *request*) an den Server und somit auch kein Logeintrag.

Auch wenn die ermittelten Server-Sessions in der Regel nicht eine 100%-ige Rekonstruktion der wahren User-Sessions darstellen, haben Server-Sessions dennoch meist eine ausreichende Qualität zur Durchführung sinnvoller Analysen des Nutzerverhaltens (Spiliopoulou u. a. 2003).

Nach dem Sessionizing folgt als nächster Teilschritt das Data Cleaning. Dieser Schritt kann auch vor dem Sessionizing durchgeführt werden, wenn es nicht darum geht, komplett Sessions zu entfernen.

Data Cleaning Mittels Data Cleaning werden alle uninteressanten Logeinträge und Sessions entfernt. Zu diesen gehören Einträge, die durch den Download von unwichtigen Dateien, wie z.B. Grafiken oder Konfigurationsdateien, erzeugt wurden. Uninteressante Sessions sind beispielsweise Administrator- und insbesondere Webrobot-Sessions. Webrobots sind Computerprogramme (z. B. Suchmaschinen), die eigenständig zwecks Informationssuche auf Websites zugreifen. Durch Webrobotaktivität erzeugte Logeinträge müssen entfernt werden, da sie nicht dem wahren Nutzerverhalten entsprechen und daher das Bild des wahren Nutzerverhaltens verzerrten könnten. Die meisten („ethischen“) Webrobots offenbaren ihre Identität und können daher im Log leicht erkannt werden. Einige („unethische“) sind nur durch Analyse ihres Verhaltens identifizierbar (Tan u. Kumar 2000, 2002).

Structure and Content Preprocessing Der letzte Teilschritt der Datenselektions- und -aufbereitungsphase besteht in einer zumindest minimalen Auswertung von Inhalten und Strukturen der betrachteten Website.

Beim Content Preprocessing werden Seiten, z. B. ihrer Funktion oder ihrem Inhalt entsprechend, in verschiedene Webdokumenttypen wie Homepage, Navigations-Seite, Content-Seite usw. klassifiziert (z. B. Berendt u. Spiliopoulou 2000; Cooley u. a. 1999; Pirolli u. a. 1996). Die anschließende Analyse der Websitestruktur (z. B. Spertus 1997) macht dann eine sinnvolle Betrachtung von Besucherpfaden erst möglich. Unter anderem können die hier erfassten Informationen zusammen mit der Auswertung des Nutzerverhaltens hilfreich bei der Identifikation von Schwachstellen in Websites sein (z. B. Spiliopoulou u. Berendt 2002; Sullivan 1997).

Mit der Inhalts- und Strukturanalyse ist die Datenselektions- und -aufbereitungsphase abgeschlossen. Die vorverarbeiteten Daten können jetzt in der nächsten Phase mit verschiedenen Data Mining Methoden auf interessante Muster untersucht werden.

Musterextraktion Übliche Data Mining Methoden zur Musterextraktion sind die Klassifikation, die Clusteranalyse, das Assoziationsregelverfahren und die Ge-

neralisierung (Ester u. Sander 2000). Zu Erläuterungen dieser Methoden siehe Tabelle 3.7.

Ein weiteres insbesondere für das Web Usage Mining interessantes Verfahren ist die Analyse sequentieller Muster (Agrawal u. Srikant 1995; Srikant u. Agrawal 1996). Mit diesem Verfahren können komplett Klickpfade von Besuchern einer Website rekonstruiert und analysiert werden (z.B. Mobasher u. a. 1996; Spiliopoulou 2000).

Tabelle 3.7: Data Mining Methoden

Methode	Beschreibung
Klassifikation	Die untersuchten Objekte werden ihren Eigenschaften entsprechend vordefinierten Klassen zugeordnet. Die Zuordnung erfolgt auf Basis der Anzahl und Stärke der Übereinstimmungen zwischen Eigenschaften eines Objekts und einer definierten Klasse.
Clusteranalyse	Objekte werden anhand ihrer Eigenschaften gruppiert bzw. <i>geclustert</i> . Die Übereinstimmung der Eigenschaften von Objekten innerhalb eines Clusters soll dabei möglichst groß und die zwischen zwei Clustern möglichst gering sein.
Assoziationsregelverfahren	Es wird analysiert, mit welcher Wahrscheinlichkeit zwei oder mehr Objekte der betrachteten Menge in einer Transaktion (z. B. eine Session) zusammen vorkommen. Beispiel: Wenn ein Besucher die Homepage anklickt, dann wird er mit einer Wahrscheinlichkeit von 30% auch die News-Seite anklicken.
Sequenzanalyse	Die Sequenzanalyse ist eine Erweiterung des Assoziationsregelverfahrens. Zusätzlich zur Information über zusammen vorkommende Objekte, werden auch Informationen zu deren Reihenfolge extrahiert. Beispiel: Wenn ein Besucher die Homepage anklickt, dann wird er mit einer Wahrscheinlichkeit von 30% unmittelbar darauf die News-Seite anklicken.
Generalisierung	Einzelne Objekte werden klassifiziert bzw. auf allgemeinere Konzepte abgebildet. Die Analyse erfolgt dann auf der Ebene dieser Konzepte (siehe Einführung zu Konzeptthierarchien in Abschnitt 2.2 auf Seite 11), d. h. an Stelle von konkreten Daten werden Zusammenhänge zwischen aggregierte Daten betrachtet.

Welche Methode der Musterextraktion eingesetzt wird, hängt vom Ziel der Untersuchung ab. Es lassen sich fünf Anwendungsgebiete unterscheiden (Abbildung 3.11): (1) Personalisierung, (2) Websiteanpassung, (3) Systemverbesserung, (4) Generierung von BI-Informationen und (5) Charakterisierung der Nutzung (Srivastava u. a. 2000).

- **Personalisierung:** Ziel der Personalisierung ist es, Inhalte, Formen, Funktionen und Strukturen einer Website an nutzerspezifische Bedürfnisse anzupassen. Der Ansatz basiert auf der Bildung von Nutzergruppen auf Basis

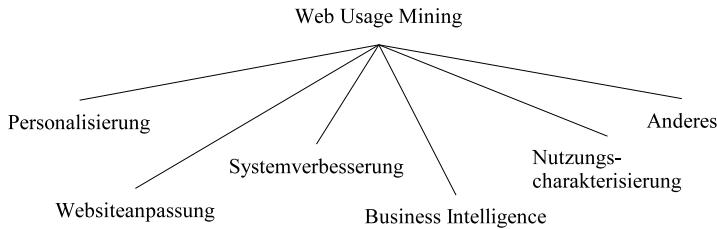


Abbildung 3.11: Anwendungsgebiete des Web Usage Mining

der verschiedenen vorhandenen Verhaltensmuster unter den Nutzern. Neue Besucher einer Website werden dann anhand ihres Verhaltens einer Gruppe zugeordnet und die in dieser Gruppe bevorzugten Inhalte, Layouts, Farben usw. werden ihnen präsentiert. Bestehende Ansätze nutzen dazu unter anderem Clusteranalysen (z. B. Flesca u. a. 2005; Fu u. a. 1999; Mobasher u. a. 2002), Klassifikationsverfahren (z. B. Catledge u. Pitkow 1995) und Sequenzanalysen (z. B. Spiliopoulou u. a. 2002). Auch statistische Verfahren, die nicht zu den Data Mining Methoden zählen, werden dafür eingesetzt (Dhyani u. a. 2002).

- **Websiteanpassung:** Ziel der Websiteanpassung ist es, Inhalte, Formen, Funktionen und Strukturen zu Verbessern. Dies kann unter anderem durch Verbesserung der Navigierbarkeit sowie Identifikation von toten Bereichen und den dafür möglichen Ursachen erfolgen. Die Sequenzanalyse eignet sich hierfür besonders, da Nutzungsmuster sehr detailliert ausgewertet und Schwachstellen präzise lokalisiert werden können (z. B. Spiliopoulou u. Berrendt 2001; Spiliopoulou u. Pohle 2001). Auch statistische Verfahren, die nicht zu den Data Mining Methoden zählen, werden dafür eingesetzt. Vorgeschlagene Ansätze basieren z. B. auf einer simplen Auswertung von Zugriffszahlen (Sullivan 1997) oder auf einer Klassifikation von Webdokumenten auf Basis ihrer Nutzung (Fu u. a. 2001).
- **Systemverbesserung:** Performance, Sicherheit und Robustheit sind Faktoren, die die Basissystemqualität einer Website charakterisieren und Einfluss auf die Nutzerzufriedenheit ausüben können. Gegenstand der Systemverbesserung ist es, Schwachstellen durch Auswertung von Zugriffszahlen und -mustern zu erkennen und zu verbessern (z. B. Cohen u. a. 1998).
- **Business Intelligence:** BI kann definiert werden als

[...] analytische[r] Prozess, der - fragmentierte - Unternehmens- und Wettbewerbsdaten in handlungsgerichtetes Wissen über die Fähigkeiten, Positionen, Handlungen und Ziele der betrachteten internen oder externen Handlungsfelder (Akteure und Prozesse) transformiert. (Grothe u. Gentsch 2000, S. 19)

Beispiel für die Generierung von BI-Informationen im Webkontext ist die Konsolidierung von Logdaten mit Marketingdaten in einem Web-Logdaten-Hypercube zur Extraktion von *Marketing Intelligence* für eCommerce (Büchner u. Mulvenna 1998).

- **Charakterisierung von Webnutzung:** Bei der Charakterisierung von Web-Nutzungsverhalten wird in die Betrachtung des Navigationsverhaltens eines Nutzers auch sein Bedienverhalten der Browsersoftware (z.B. Betätigung des „Zurück“-Buttons, Speichern eines Links) einbezogen und ausgewertet. So lassen sich Zusammenhänge zwischen Nutzer, Browserbedienung und Navigationsstrategie analysieren (z.B. Catledge u. Pitkow 1995).

Websites können aufgrund ihrer Größe und/oder durch dynamisch erzeugte Inhalte sehr viele verschiedene Seiten enthalten. Durch die Vielzahl der sich daraus ergebenden möglichen Kombinationen von Seitenabruften wird eine Analyse auf der Ebene konkreter Seiten in der Regel wenig aussagekräftige Resultate liefern. Mittels Generalisierung können hingegen aussagekräftige Zusammenhänge abgeleitet werden. Seiten können dafür z.B. nach Inhalt oder Funktion klassifiziert werden (Pohle u. Spiliopoulou 2002). Die Auswertung generalisierter Webnutzungsmuster ist dann prinzipiell mit jeder Data Mining-Methode möglich, wie z.B. die Cluster- (Fu u. a. 2000) oder die Sequenzanalyse (Pohle u. Spiliopoulou 2002; Spiliopoulou u. Berendt 2001; Spiliopoulou u. Pohle 2001).

Zur Erfassung von zeitbezogenen Mustern, Charakteristiken, Trends usw. können die beschriebenen Ansätze auch im Rahmen von Zeitreihenanalysen eingesetzt werden (z.B. Moe u. Fader 2004; Zaïane u. a. 1998).

Zur Erweiterung des Bewertungsinstrumentariums oder als Alternative zu den rein quantitativen Ansätzen können Nutzungsmuster visualisiert werden (z.B. Berendt 2002; Lee u. a. 2000a, 2001). Diese Art der Darstellung ermöglicht eine einfache und schnelle Erfassung von Strukturen in den Daten.

Musterinterpretation Die letzte Phase des Web Usage Mining Prozesses besteht in der Interpretation der extrahierten Muster. Dies erfolgt auf Basis von Kontextinformationen und den Auswertungsergebnissen zu Inhalten, Strukturen und Nutzerverhalten. Erst in dieser Phase werden aus den extrahierten Daten Nutzen bringende Informationen. Der Web Usage Mining Prozess ist damit abgeschlossen.

Statistische Verfahren

Neben Data Mining werden auch (simple) statistische Verfahren zur Analyse von Nutzerverhalten eingesetzt.

Auswertungen von Häufigkeiten, wie z.B. die Anzahl an Zugriffen pro Zeiteinheit (Stunde, Tag, Monat usw.) oder die Anzahl an Zugriffen pro Seite (siehe Hightower u. a. 1998; Sullivan 1997), gehören zu den Basisfunktionen gängiger Loganalyse-Tools (z.B. Analog, AWStats, Webalizer). Die Zahlen liefern einen

schnellen Überblick zu Zugriffen auf eine Website. Zusammenhänge zwischen einzelnen Requests werden im Gegensatz zu den Data Mining-basierten Ansätzen allerdings nicht erfasst.

Über die reine Erfassung von Häufigkeiten hinaus gehen Ansätze, die aus den Daten interessante Kennzahlen (engl.: *metrics*), wie z.B. Kundenaquisitions- oder Kundenkonversionsraten, ermitteln (z.B. Berthon u.a. 1996; Cutler u. Sterne 2000). Diese können unter anderem zur Analyse des Lebenszyklus von Kundenbeziehungen eingesetzt werden (Cutler u. Sterne 2000).

Neben der Ableitung von Kennzahlen kann mittels statistischer Auswertung auch zukünftiges Nutzerverhalten prognostiziert werden (Dhyani u.a. 2002).

Webseiteninhalte können ebenfalls mit statistischen Methoden analysiert werden. So kann z.B. die Relevanz einer Webseite in Bezug auf ein bestimmtes Thema an Hand der Häufigkeit von im Text vorkommenden Schlüsselbegriffen oder auf Basis der Anzahl anderer Seiten mit Verlinkungen auf diese Seite ermittelt werden (Dhyani u.a. 2002).

Datenschutzaspekte

Logdaten können sensible Informationen wie eindeutige Benutzernamen oder IP-Adressen enthalten, an Hand derer Nutzer identifizierbar sind. Aus diesem Grund ist ein sorgsamer Umgang mit diesen Daten wichtig, so dass Sicherheit und Privatsphäre der Nutzer gewahrt bleiben und dennoch möglichst viele nützliche Informationen extrahiert werden können.

Grundsätzlich wird eine Diskrepanz zwischen der Bereitschaft eines Users zur Preisgabe persönlicher Informationen und dem Verlangen des Website-Anbieters nach möglichst umfangreichen personenspezifischen Daten vorliegen. Der Nutzer hat das Recht und in der Regel auch das Bedürfnis nach Privatsphäre und der Sicherheit seiner Daten. Auf der anderen Seite braucht ein Website-Anbieter möglichst umfangreiche Informationen, um seine Website optimal an die Bedürfnisse der Nutzer anpassen zu können.

Für einen guten Mittelweg können entsprechende Richtlinien wie die *European directive on privacy and electronic communications* (DIRECTIVE 2002/58/EC) als Basis zum Umgang mit Logdaten herangezogen werden.

Vor- und Nachteile der Logdaten-Auswertung

Zum Schluss dieses Abschnitts wird eine kurze, zusammenfassende Übersicht zu den Vor- und Nachteilen der Auswertung von Nutzerverhalten auf Basis von Logdaten gegeben.

Vorteile sind unter anderem:

- Alle im betrachteten Untersuchungszeitraum festgehaltenen Nutzeraktivitäten können in die Auswertung einbezogen werden. Man ist also nicht, wie z.B. bei Befragungen üblich, darauf angewiesen, eine repräsentative Stichprobe der Grundgesamtheit zu haben.

- Die Auswertung übt keinen Einfluss (wie es z. B. bei Experimenten der Fall sein kann) auf das Nutzerverhalten aus, da sie vom Nutzer nicht wahrgenommen wird.
- Der Analyseaufwand ist im Vergleich zu z. B. Umfragen oder Experimenten gering. Eine regelmäßige Durchführung ist daher ohne weiteres möglich.

Als Nachteile lassen sich unter anderem folgende Punkte aufführen:

- Die Qualität der Ergebnisse hängt entscheidend von der Vorverarbeitung ab. Nutzersessions können aufgrund der oben beschriebenen Problematik in vielen Fällen nicht vollständig rekonstruiert werden, was zu leicht verzerrten oder im Extremfall zu falschen Ergebnissen führen kann. Nicht identifizierte Webrobotsessions und andere uninteressante Sessions, die nicht aus dem Log entfernt wurden, können ein falsches Bild des Nutzerverhaltens wiedergeben.
- Auswertungsergebnisse sind nicht immer eindeutig interpretierbar. So z. B. kann eine große Zahl an Page-Views während einer Session auf einen sehr interessierten oder aber auf einen desorientierten Nutzer hinweisen. Die Be- trachtung einer Seite über einen längeren Zeitraum kann darauf hinweisen, dass der Inhalt für den User sehr interessant ist oder aber, dass dieser sich kurzzeitig vom Rechner entfernt hat oder dort zeitgleich einer anderen Beschäftigung nachgeht (Sullivan 1997). In solchen Fällen ist zusätzliches Hintergrund- und Kontextwissen hilfreich und notwendig für eine Interpretation der Ergebnisse.
- Es ist nur das Nutzerverhalten erfassbar. Über die Nutzermeinungen – welche eine weitere wichtige Informationsquelle für die Erfolgseinschätzung sind – können auf Basis der Interpretation des Verhaltens nur Vermutungen ange- stellt werden. Genaue Aussagen lassen sich nicht ableiten.

3.3.3 Expertenbasierte Analyse der Websitequalität

Der dritte verbreitete Ansatz neben der Analyse von Nutzermeinungen und Nutzerverhalten ist die expertenbasierte Analyse von Websitequalität. Gegenstand ist die Untersuchung und Bewertung einer Website durch einen „Experten“ anhand subjektiv wahrgenommener sowie objektiv erfassbarer Merkmale. Die Analyse basiert auf Erfahrungen und Wissen des Experten und/oder Ergebnissen von Forschungsarbeiten⁶ über Zusammenhänge verschiedener Websitecharakteristiken und Effekten der Nutzung, wie z. B. dem Einfluss von Informationsqualität auf die Nutzerzufriedenheit.

Die expertenbasierte Analyse wird häufig im Rahmen einer Usability-Bewertung⁷ durchgeführt. Dort spricht man auch von „Usability-Inspektion“ (Ivory u. Hearst

⁶Siehe dazu die aufgeführten Arbeiten in Abschnitt 3.2.8.

⁷Siehe dazu die Einführung des Begriffs „Usability“ in Abschnitt 3.2.

2001). Die angewendeten Methoden können nach der Art ihrer Durchführung in manuelle und automatische Inspektionsmethoden untergliedert werden (Ivory u. Hearst 2001). Bei der manuellen Inspektion bewertet ein Experte oder ein Expertenteam mittels z. B. Fragebögen, Checklisten oder Szenarien die Qualität bzw. Usability einer Webpräsenz. Die Bewertung basiert dabei auf der individuellen Einschätzung des Evaluators oder auf objektiv messbaren Eigenschaften der betrachteten Website. Bei der automatisierten Inspektion erfolgt die Analyse ausschließlich auf Basis von Kriterien zu objektiv erfassbaren Merkmalen mittels des Einsatzes einer Softwareapplikation (Ivory u. Hearst 2001).

In beiden Ansätzen wird für die Aufstellung von Bewertungskriterien ein Regelwerk basierend auf definierten Standards, Richtlinien, Prinzipien und ähnlichem genutzt (z. B. Keeker 1997; Nielsen 2000; Nielsen u. Mack 1994). Beispiel für ein entsprechendes Regelwerk sind die 10 Usability-Design-Prinzipien von Nielsen (1994) in Tabelle 3.8.

Tabelle 3.8: Nielsens 10 Usability-Design-Prinzipien

1	<i>Visibility of system status</i>
2	<i>Match between system and the real world</i>
3	<i>User control and freedom</i>
4	<i>Consistency and standards</i>
5	<i>Error prevention</i>
6	<i>Recognition rather than recall</i>
7	<i>Flexibility and efficiency of use</i>
8	<i>Aesthetic and minimalist design</i>
9	<i>Help users recognize, diagnose, and recover from errors</i>
10	<i>Help and documentation</i>

Quelle: Nielsen (1994)

Die in Webusability-Ansätzen verwendeten Kriterien lassen sich entsprechend der Klassifikation in Abschnitt 3.2.8 in Kriterien zur Bewertung von Informationsqualität, Systemqualität und Servicequalität einer Website untergliedern. In Abhängigkeit des jeweils gewählten Kriteriums erfolgt die Bewertung manuell auf Basis subjektiver Einschätzung des Evaluators oder durch manuelle oder automatisierte Messung objektiv erfassbarer Merkmale. Die zur subjektiven Einschätzung zugrunde gelegten Kriterien entsprechen weitgehend den in Nutzerbefragungen verwendeten (siehe Abbildung 3.12 und Tabelle 3.9).

Für die manuelle Inspektion von Informationsqualität basierend auf subjektiver Einschätzung werden Kriterien, wie z. B. Aktualität, Relevanz und Exaktheit der Informationen, herangezogen (z. B. Agarwal u. Venkatesh 2002; Alexander u. Tate 1999; Huizingh 2000; Moustakis u. a. 2004; Palmer 2002; Sampson u. Manouselis 2005). Verwendete Kriterien zur manuellen Erfassung objektiv messbarer Merkmale sind unter anderem das Vorhandensein einer Sitemap, eines Impressums oder einer FAQ-Seite (z. B. Alexander u. Tate 1999; Moustakis u. a. 2004).

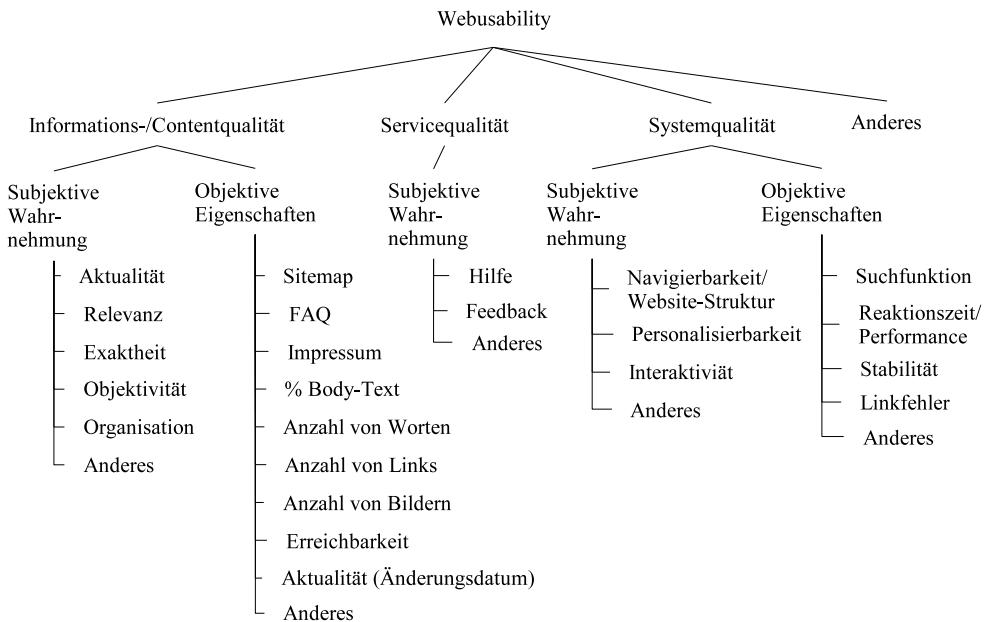


Abbildung 3.12: Aus den betrachteten Ansätzen synthetisierte Klassifikation von Bewertungskriterien zur Webusability-Inspektion

Die Auswertung anderer objektiv erfassbarer Eigenschaften, wie z. B. die Anzahl an Worten, Bildern oder Links pro Seite, manuell durchzuführen, kann bei großen Websites sehr aufwändig werden. Mit automatisierten Verfahren ist das in solchen Fällen einfacher und exakter umsetzbar (z. B. Ivory u. a. 2001; Knight u. Burn 2005; Olsina u. a. 1999; Olsina u. Rossi 2002; Palmer 2002).

Für die manuelle Inspektion der Systemqualität basierend auf subjektiver Einschätzung werden Kriterien, wie z. B. Navigierbarkeit, Struktur, Personalisierbarkeit oder Interaktivität der Website, herangezogen (z. B. Agarwal u. Venkatesh 2002; Huizingh 2000; Moustakis u. a. 2004; Olsina u. a. 1999; Olsina u. Rossi 2002; Palmer 2002). Objektiv erfassbare Merkmale sind unter anderem das Vorhandensein einer Suchfunktion, die Systemreaktionszeit, Browserkompatibilität, Anzahl kaputter Links oder Systemstabilität (z. B. Huizingh 2000; Moustakis u. a. 2004; Olsina u. a. 1999; Olsina u. Rossi 2002). Diese Eigenschaften können sowohl manuell als auch automatisiert erfasst werden. Kriterien wie Systemreaktionszeit oder Anzahl kaputter Links eignen sich aufgrund des erhöhten Messaufwands eher für eine automatisierte Auswertung (z. B. Olsina u. a. 1999; Olsina u. Rossi 2002; Palmer 2002).

Ansprechbarkeit bzw. Entgegenkommen (engl.: *responsiveness*) eines Systems wird als ein Merkmal für IS-Servicequalität definiert (s. Abschnitt 3.2.8). Dieses kann z. B. durch das Kriterium „Feedback“ (des Website-Systems) operationalisiert werden (Palmer 2002). Ein weiteres in diese Kategorie fallendes Kriterium wäre

Tabelle 3.9: Ansätze zur Usability-Inspektion von Websites

Autor(en)	IQ/CQ		SQ		SRQ		A
	Manuell	Auto	Manuell	Auto	Manuell	Auto	
Agarwal u. Venkatesh (2002)	✓			✓			✓
Alexander u. Tate (1999)	✓						
Huizingh (2000)	✓			✓			
Ivory u. a. (2001)			✓				
Knight u. Burn (2005)			✓				
Moustakakis u. a. (2004)	✓			✓			✓
Olsina u. a. (1999)	✓		✓	✓		✓	
Olsina u. Rossi (2002)			✓			✓	
Palmer (2002)	✓		✓	✓	✓	✓	✓
Sampson u. Manouselis (2005)	✓			✓			✓

✓ vorhanden

IQ/CQ Informations-/Contentqualität SQ Systemqualität
SRQ Servicequalität A Anderes

die Hilfefunktionalität der Website (Olsina u. a. 1999). Auch wenn diese Attribute nicht immer explizit als Kriterien zur Erfassung von Website-Serviceequalität definiert werden (z.B. Agarwal u. Venkatesh 2002; Palmer 2002; Olsina u. a. 1999), können sie als solche interpretiert werden (z.B. Wang u. Tang 2003).

Ein großer Vorteil der expertenbasierten Inspektion ist, dass keine Zusammenarbeit mit anderen Stakeholdern nötig ist. Es werden auch keine speziellen Datenquellen (wie bei Umfragen oder Logdatenanalysen) benötigt. Die Bewertung erfolgt allein auf Basis einer Betrachtung bzw. Inspektion der Website. Mittels Automatisierung lässt sich dieser Prozess sogar noch vereinfachen und beschleunigen. Manuelle Untersuchungen sind dagegen unter Umständen sehr aufwändig. Außerdem ist nicht sicher, ob dabei alle Usability-Probleme aufgedeckt werden. So berichten Nielsen u. Mack (1994, S. 14) von einer Trefferquote zwischen 30% und 50%. Daneben ist unklar, ob die von Experten gefundenen „Probleme“ auch tatsächlich welche in den Augen der Nutzer darstellen. Ein weiterer Nachteil ist die zum Teil sehr abstrakte Formulierung von Richtlinien, Prinzipien usw. Es ist nicht immer eindeutig, wie diese umgesetzt werden sollen (z.B. „Recognition rather than recall“, siehe Tabelle 3.8).

3.4 Zusammenfassung

In diesem Kapitel wurde eine Einführung zu verschiedenen bestehenden Evaluationsansätzen für IS gegeben und basierend auf diesen ein Rahmenmodell zur Konzeption eines IS- und Kontextspezifischen Evaluationsansatzes entwickelt. Bezug nehmend auf dieses Rahmenmodell wurden verbreitete Ansätze zur Analyse von Websites vorgestellt.

IS sind Systeme die sowohl aus technischen als auch aus humanen Komponen-

ten (mindestens den Nutzern) bestehen. Eine Evaluation ist daher sowohl mittels quantitativer als auch qualitativer Ansätze möglich. Daneben sind auch Kombinationen, d. h. pluralistische Ansätze, denkbar. Diese Ausgangssituation hat zu einer Vielzahl unterschiedlicher Vorschläge zur Bewertung von IS-Erfolg geführt. Welcher von diesen die beste Lösung beinhaltet, ist bis dato nicht geklärt. Fest steht, dass trotz einer ansatzweise vorhandenen kumulativen IS-Forschungstradition, diese noch stark fragmentiert ist. Es gibt daher keinen Bewertungsansatz, der den anderen deutlich überlegen ist. Vielmehr ist aus dem bisherigen Stand der Forschung der Schluss zu ziehen, dass für ein spezifisches IS in einem spezifischen Kontext ein aus den bestehenden Ansätzen maßgeschneidertes Evaluationskonzept die derzeit beste Lösung ist. Ein entsprechendes Rahmenmodell, welches bei der Konzipierung Hilfestellung leisten soll, wurde hier entwickelt.

Drei verbreitete Vorgehensweisen zur Evaluation von Websites wurden, unter Bezugnahme auf dieses Rahmenmodell und drei wichtige Modelle zu relevanten Variablen von IS-Erfolg, vorgestellt: die Analyse von Nutzermeinungen, die Analyse von Nutzerverhalten und die expertenbasierte Analyse von Websitequalität.

Kapitel 4

Multikriterielle Entscheidungsanalyse (MCDA)

Vorliegendes Kapitel gibt eine Einführung zu Ansätzen der Multikriteriellen Entscheidungsanalyse.

In Abschnitt 4.1 wird das Grundmodell Multikriterieller Entscheidungsanalyse vorgestellt. Abschnitt 4.2 gibt einen Überblick zu wichtigen Verfahren. In Abschnitt 4.3 wird detailliert auf das in dieser Arbeit verwendete Verfahren – die Nutzwertanalyse – eingegangen. Abschnitt 4.4 fasst das Kapitel zusammen.

4.1 Grundmodell

Eine Entscheidung ist „eine bewusste Wahl zwischen Alternativen oder zwischen mehreren unterschiedlichen Varianten anhand bestimmter Präferenzen von einem oder mehreren Entscheidungsträgern“ (Wikipedia.org 2008a). Ein Entscheidungsproblem lässt sich anhand von drei wichtigen Merkmalen charakterisieren (Schneeweiss 1991, S. 93):

- A)** die Anzahl der betrachteten Ziele (Ein- oder Mehrzielentscheidungsproblem),
- B)** die Anzahl der beteiligten Entscheider (Ein- oder Mehrpersonenentscheidungsproblem),
- C)** der Bestimmtheitsgrad der Entscheidung (Sicherheit, Risiko, Ungewissheit, Unschärfe).

Zu (A): Jeder Entscheidungsprozess basiert auf einem Hauptkriterium bzw. einem so genannten *Oberziel*, auf dessen Grundlage die Entscheidungsfindung vollzogen wird. In der Regel ist dieses Oberziel nur eine Formulierung einer abstrakten Zielvorstellung, wie z.B. „IS-Erfolg“, welche nicht direkt erfass- und messbar ist. Der Entscheidungsprozess beginnt dann damit, das formulierte Oberziel durch Operationalisierung in messbare Subziele bzw. Zielkriterien greifbar zu machen. Die Operationalisierung kann aus mehreren Stufen bestehen, was zu einer Zielhierarchie, bestehend aus Oberziel, Zwischenzielen und letztlich empirisch erfassbaren Unterzielen, führt (siehe Abbildung 4.1). Drei wichtige Bedingungen sind hierbei zu beachten (Schneeweiss 1991, S. 65 f.):

1. **Operationalität:** Ergebnis der Operationalisierung müssen Subziele bzw. Zielkriterien mit mindestens ordinalem Messniveau sein, da sonst keine Entscheidungsfindung möglich ist.
2. **(Ziel-)Vollständigkeit:** Für ein betrachtetes Ziel sollten alle relevanten Subziele berücksichtigt werden, da es sonst zu einer fehlerhaften Entscheidungsfindung kommen kann.
3. **Redundanzfreiheit:** Ziele auf der gleichen Operationalisierungsstufe sollten sich jeweils auf unterschiedliche Aspekte beziehen. Wenn systematische Beziehungen (Bamberg u. Coenenberg 2002, S. 52 f.) zwischen Zielen herrschen, dann ist nicht immer der Einbezug aller Ziele notwendig. Bei z.B. wechselseitiger Abhängigkeit im Realisierungsgrad zweier Ziele über den gesamten Ergebnisbereich (symmetrische Komplementarität), braucht nur eines in den Entscheidungsprozess einzbezogen zu werden.

MCDA bezieht sich immer auf ein Mehrzielentscheidungsproblem. Der Einsatz entsprechender Verfahren ist in jenen Fällen notwendig bzw. sinnvoll, in denen unterschiedliche Alternativen bezüglich unterschiedlicher Zielvorstellungen optimal

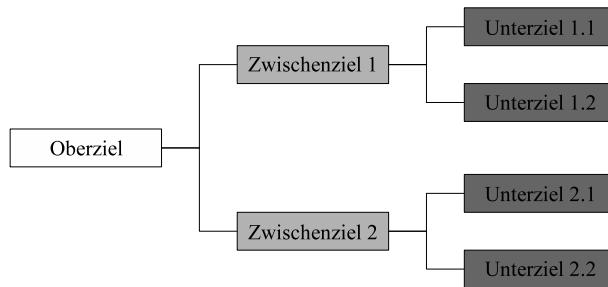


Abbildung 4.1: Beispiel einer dreistufigen Zielhierarchie

sind und der Entscheider nicht aus dem „Stegreif“ die beste zu wählen vermag. Im Rahmen solcher Entscheidungssituationen kommt es oft vor, dass Zielkriterien konfliktieren (z. B. hohe Fahrgeschwindigkeit und geringer Benzinverbrauch bei einem Auto) und/oder auf unterschiedlichen Maßeinheiten (z. B. Km/h und Liter) basieren. Die Multikriterielle Entscheidungsanalyse zielt darauf ab, in solchen Situationen unter Berücksichtigung aller Zielkriterien und der Präferenzvorstellungen des Entscheidungsträgers, auf systematische Weise, die beste oder die Menge der besser geeigneten Alternativen aus allen verfügbaren zu identifizieren.

Zu (B) und (C): Im Weiteren werden hier nur Einpersonen-Mehrziel-Entscheidungsprobleme unter Sicherheit behandelt, da diese Konstellation die Ausgangslage für das in Kapitel 5 entwickelte Bewertungsmodell darstellt. Allgemeine Einführungen zur Thematik sowie zu anderen Ausprägungen von Entscheidungsproblemen finden sich unter anderem in Bamberg u. Coenenberg (2002); Figueira u. a. (2005); Schneeweiss (1991).

Auf Basis der Entscheidungstheorie und dem hier betrachteten Entscheidungsproblem (charakterisiert in den Punkten (A), (B) und (C)) wird das Grundmodell zur Entscheidungsanalyse wie folgt spezifiziert:

Gegeben ist eine Menge $A = \{a_1, a_2, \dots, a_n\}$ an Alternativen, zu welchen mittels der betrachteten Menge $C = \{c_1, c_2, \dots, c_m\}$ an Kriterien Präferenzentscheidungen getroffen werden sollen. In diesem Zusammenhang spricht man bei Zielkriterien auch von *Attributen* bzw. *Zielattributen*. Ein Zielattribut c_j (z. B. Fahrgeschwindigkeit eines Autos) hat verschiedene Ausprägungen $x \in X$ mit X als der Menge aller seiner Ausprägungen. Diese wird als „Ausprägungsbereich“ bzw. „Domäne“ bezeichnet. Die Menge aller Attribute bildet auf Basis ihrer jeweils konkreten Ausprägungen einen Vektor, der die Charakteristiken einer Alternative definiert. Das entstehende Entscheidungsfeld zu allen betrachteten Alternativen lässt sich in einer Ergebnismatrix darstellen (siehe Abbildung 4.2). Für eine Entscheidungsfindung ist die Menge der Alternativen anhand ihrer Ergebnisausprägungen in eine Besser-Schlechter-Reihenfolge auf Basis der Präferenzinformationen des Entscheiders zu bringen.

Für Einzielentscheidungen bzw. im Ein-Attribut-Fall gilt als Voraussetzung für eine Entscheidungsfindung, dass das Attribut mindestens ordinale Messniveau hat

	c_1	c_2	\cdots	c_j	\cdots	c_m
a_1	x_{11}	x_{12}		\cdots		x_{1m}
\vdots	\vdots	\vdots		\vdots		\vdots
$E = a_i$	\cdots	\cdots	\cdots	x_{ij}	\cdots	x_{im}
\vdots	\vdots	\vdots		\vdots		\vdots
a_n	x_{n1}	x_{n2}		\cdots		x_{nm}

Abbildung 4.2: Beispielschema einer Ergebnismatrix E

und der Entscheider in der Lage ist, die Alternativen anhand der Ausprägungen dieses Attributs in eine zumindest schwache Präferenzordnung „ \lesssim “ zu bringen.

Definition 1 (Vollständigkeit). Vollständigkeit ist gegeben, wenn es keine zwei Alternativen gibt, die nicht auf Basis ihrer Attributausprägungen miteinander vergleichbar sind, d. h. für zwei beliebige Alternativen a' und a'' muss entweder $a' \lesssim a''$ oder $a' \gtrsim a''$ oder $a' \sim a''$ gelten.

Vollständigkeit verlangt, dass der Entscheidungsträger in der Lage ist, Präferenzentscheidungen zwischen zwei Alternativen treffen zu können.

Definition 2 (Transitivität). Transitivität liegt vor, wenn für alle Alternativen auf Basis ihrer Attributausprägungen gilt:

$$\begin{aligned} &\text{Aus } a' \lesssim a'' \text{ und } a'' \lesssim a''' \text{ folgt } a' \lesssim a''' \\ &\quad \text{und} \\ &\text{aus } a' \lesssim a'' \text{ und } a'' \sim a''' \text{ folgt } a' \lesssim a''' \\ &\quad \text{und} \\ &\text{aus } a' \sim a'' \text{ und } a'' \sim a''' \text{ folgt } a' \sim a'''. \end{aligned}$$

Transitivität als Präferenzgraph dargestellt, bedeutet, dass keine echten oder unechten Zyklen zugelassen sind.

Theorem 1 (Existenz einer Wertfunktion). Im Falle abzählbar vieler Alternativen und bei Vollständigkeit und Transitivität kann eine Relation numerisch durch eine Funktion $v : A \rightarrow \mathbb{R}$ dargestellt werden, die die Ordnungsrelation „ \succ, \sim “ in die numerische Beziehung „ $>, =$ “ überträgt, d. h. lassen sich die Attributausprägungen zumindest schwach ordnen, so existiert eine Wertfunktion $v : A \rightarrow \mathbb{R}$ für die gilt: $v(a') \geq v(a'') \Leftrightarrow a' \lesssim a''$. (French 1988)

Die Wertfunktion ist bis auf streng monoton wachsende Transformationen festgelegt. Für endlich viele (n) Alternativen stellt $v(a)$ eine Rangzahl dar, die sich durch $n(n - 1)$ Paarvergleiche ermitteln lässt.

Definition 3 (Höhenpräferenz). Die Höhenpräferenz drückt die Präferenzvorstellung des Entscheidungsträgers zur Anordnung der Alternativen bezüglich eines speziellen Attributes aus.

Definition 4 (Artenpräferenz). Die Artenpräferenz drückt bei Mehrzielentscheidungen die Präferenzvorstellung des Entscheidungsträgers bezüglich der relativen Wichtigkeit eines Zielattributes im Vergleich zu den anderen Attributen aus.

Bei Mehrzielentscheidungen muss der Entscheider im Gegensatz zur Einzielentscheidung seine Höhenpräferenz bezüglich aller Zielattribute ausdrücken können. Er kann dabei Attribute in ihrem Grad der Wichtigkeit unterscheiden (Artenpräferenz). Für alle Attribute gelten die Restriktionen, wie im Ein-Attribut-Fall: Mindestens ordinale Messniveau, Vollständigkeit und Transitivität.

Definition 5 (Schwache Präferenzunabhängigkeit). Schwache Präferenzunabhängigkeit zwischen zwei Alternativen bzw. Attributvektoren ist gegeben, wenn für ein Zielattribut $c_j \in C$, alle übrigen Attribute $c \in C$ ($c_j \neq c$) und für zwei beliebige Attributausprägungen $\bar{x} \neq \bar{\bar{x}}$ gilt, dass wenn

$$\begin{aligned} (\bar{x}_1, \dots, \bar{x}_{j-1}, y_j, \bar{x}_{j+1}, \dots, \bar{x}_m) &\succsim (\bar{x}_1, \dots, \bar{x}_{j-1}, z_j, \bar{x}_{j+1}, \dots, \bar{x}_m), \\ &\text{daraus} \\ (\bar{\bar{x}}_1, \dots, \bar{\bar{x}}_{j-1}, y_j, \bar{\bar{x}}_{j+1}, \dots, \bar{\bar{x}}_m) &\succsim (\bar{\bar{x}}_1, \dots, \bar{\bar{x}}_{j-1}, z_j, \bar{\bar{x}}_{j+1}, \dots, \bar{\bar{x}}_m) \end{aligned}$$

folgt. D. h. die Präferenzrelation zwischen zwei Ausprägungen y_j und z_j eines Attributes c_j ist unabhängig von den Ausprägungen der übrigen Attribute. (Schneeweiß 1991, S. 96)

Ist schwache Präferenzunabhängigkeit gegeben, existiert für jedes Zielattribut c_j eine individuelle Wertfunktion $v_j(x_j(a))$. In diesem Fall lässt sich aus der allgemeinen Ergebnismatrix eine dazu analoge Entscheidungsmatrix ableiten (Abbildung 4.3). Die Artenpräferenz wird mittels Gewichtungsfaktor w_j ($\sum_j w_j = 1$, $w_j \geq 0$) angegeben. Die anschließende Wertaggregation zur Ermittlung der Alternativenreihenfolge erfolgt auf Basis des jeweils verwendeten MCDA-Verfahrens. Eine Möglichkeit besteht in der Konstruktion einer Mehrzielpräferenzfunktion

$$\Phi(a) = f(v_1(x_1(a)), v_2(x_2(a)), v_3(x_3(a)), \dots, v_m(x_m(a))), \quad (4.1)$$

die analog zum Einzelfall eine Anordnung der Alternativen erbringt. Die beste Alternative ist dann bestimmt durch $\Phi(a^*) = \max_{a \in A} \Phi(a)$. Im einfachsten Fall kann man für f eine lineare Funktion verwenden:

$$\Phi(a) = \sum_{j=1}^m w_j \cdot v_j(x_j(a)). \quad (4.2)$$

Im folgenden Abschnitt wird eine kurze Übersicht zu wichtigen MCDA-Verfahren gegeben.

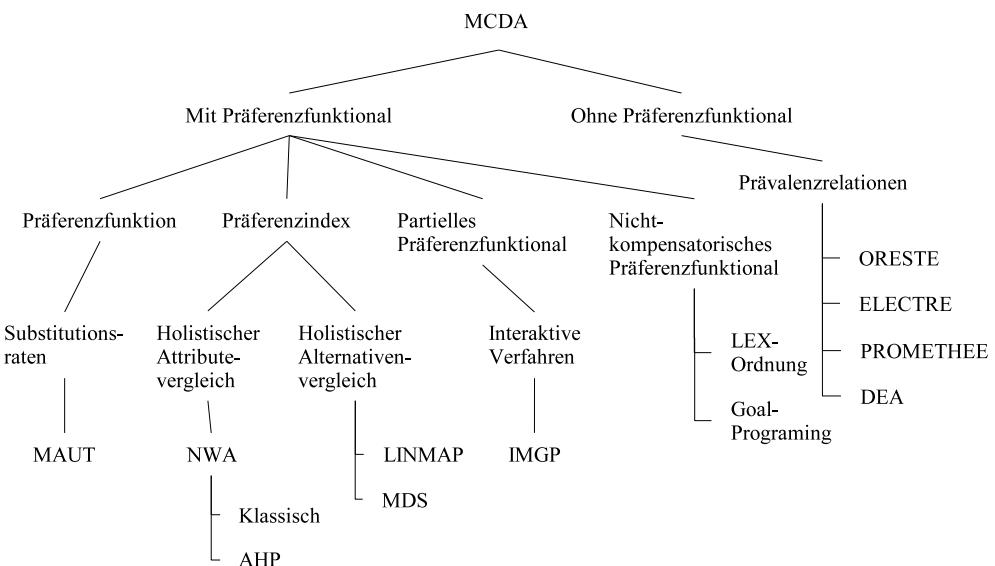
	w_1	w_2	\dots	w_j	\dots	w_m
	c_1	c_2	\dots	c_j	\dots	c_m
a_1	$v_1(x_{11})$	$v_2(x_{12})$		\dots		$v_m(x_{1m})$
\vdots	\vdots	\vdots		\vdots		\vdots
$D = a_i$	\dots	\dots	\dots	$v_j(x_{ij})$	\dots	$v_m(x_{im})$
\vdots	\vdots	\vdots		\vdots		\vdots
a_n	$v_1(x_{n1})$	$v_2(x_{n2})$		\dots		$v_m(x_{nm})$

Abbildung 4.3: Beispieldiagramm einer Entscheidungsmatrix D

4.2 Verfahren

Die hier vorgenommene Kategorisierung (Abbildung 4.4) basiert auf der Darstellung in Schneeweiss (1991, S. 291 ff.). Detaillierte Einführungen zu diesen und weiteren Verfahren finden sich unter anderem in Bamberg u. Coenenberg (2002, S. 61 ff.); Figueira u. a. (2005); Schneeweiss (1991, S. 291 ff.).

Mehrzielverfahren können zunächst in Verfahren mit und ohne Präferenzfunktional unterschieden werden. Verfahren mit Präferenzfunktional können nur dann angewendet werden, wenn der Entscheidungsträger in der Lage ist, eine zumindest schwache Präferenzordnung der Alternativen vorzunehmen. Bei Verfahren ohne Präferenzfunktional besteht diese Notwendigkeit nicht. Solche Verfahren eignen sich daher besonders für den Fall, dass der Entscheider sich nicht in der Lage sieht, sämtliche Alternativen transitiv anzugeben.



Quelle: Schneeweiss (1991, S. 293)

Abbildung 4.4: Verfahren der Mehrzielanalyse

4.2.1 Verfahren mit Präferenzfunktional

Die Verfahren lassen sich entsprechend der Ausprägung ihres Präferenzfunktionalen in Verfahren mit einer Präferenzfunktion, mit einem Präferenzindex, mit einem partiellen Präferenzfunktional und mit einem nicht-kompensatorischen Präferenzfunktional unterscheiden.

Die höchste Präferenzinformation erfordern Verfahren, die auf der Bildung einer Präferenzfunktion basieren. Dazu muss der Entscheider in der Lage sein, Substitutionsraten zwischen den Ausprägungen unterschiedlicher Attribute angeben zu können. Die Substitutionsbedingung erfordert kardinales Messniveau der Attribute. Aus Sicht des Entscheidungsträgers liegt starke Präferenzunabhängigkeit vor. Die Ergebnisse sind unabhängig von der Menge der betrachteten Alternativen, d. h. neue Alternativen können in die bestehende Ordnung eingebettet werden, ohne diese zu beeinflussen. Zu dieser Klasse von Verfahren gehört z. B. die Multi-Attributive Nutzentheorie (engl.: *Multi-Attribute Utility Theory (MAUT)*).

Bei Verfahren mit Präferenzindex braucht der Entscheidungsträger keine Substitutionsraten anzugeben, sondern muss lediglich eine mindestens schwache Ordnung der Alternativen vornehmen können. Vertreter dieser Klasse sind Verfahren der klassischen Nutzwertanalyse (NWA). Diese zeichnen sich dadurch aus, dass im Gegensatz zur MAUT Wertfunktion und Gewichtung getrennt bestimmt werden und die Gewichtung durch ganzheitliche Betrachtung (holistisch) aller Zielkriterien erfolgt.

Eine spezielle Variante der Nutzwertanalyse ist der *Analytic Hierarchy Process (AHP)*. Wie bei der NWA üblich, werden die Gewichte holistisch und unabhängig von den Wertfunktionen ermittelt. Der Unterschied zur klassischen NWA liegt darin, dass dies durch paarweise Vergleiche der Attribute geschieht. Diese Besonderheit ermöglicht dem AHP die Bestimmung von zu konsistenten Situationen führenden Wertfunktionen, auch für den Fall, dass die Angaben des Entscheidungsträgers inkonsistent sind.

Eine mittels Verfahren der NWA ermittelte Rangfolge basiert auf der konkret vorliegenden Menge der betrachteten Alternativen. Eine neu hinzugekommene Alternative kann diese Ordnung daher durcheinander bringen, obwohl sich die Präferenzvorstellung des Entscheidungsträgers nicht verändert hat. Diese Eigenschaft der NWA-Verfahren bezeichnet man als *Instabilität* bezüglich Änderungen in der betrachteten Alternativenmenge.

Globale Kompromissverfahren führen ebenfalls zu einem Präferenzindex. Im Gegensatz zur NWA werden nicht lokale Attributvergleiche sondern globale Alternativenvergleiche unter gleichzeitiger Betrachtung aller Attribute vorgenommen. Intransitivitäten sind hier zugelassen. Die Bestimmung der Gewichte erfolgt in einem Optimierungsprozess durch das Verfahren selbst, mit dem Ziel die ursprünglich artikulierten Präferenzen so gut wie möglich zu repräsentieren. Zu den globalen Kompromissverfahren gehören unter anderem das *LINear Programming techniques for Multidimensional Analysis of Preference (LINMAP)* und die Multidimensionale Skalierung (MDS).

Auf der Nutzung eines partiellen Präferenzfunktional basieren interaktive Verfahren. Bei diesen Verfahren wird das Präferenzfunktional nicht von vornherein vollständig festgelegt, sondern nur der Ausschnitt ermittelt, der die Auswahl der besten Alternative(n) gestattet. Die Vorgehensweise ist oft so, dass in einem iterativen Prozess zunächst sämtliche effiziente Alternativen gewählt werden und dann über diese Menge das als existent vorausgesetzte Präferenzfunktional konstruiert wird. Eine Alternativenmenge muss dabei nicht explizit vorgegeben sein, sondern ergibt sich implizit aus dem Restriktionenraum. Damit einhergehend unterliegt hier wie bei MAUT die Annahme eines nicht begrenzten Alternativenraums. Beispiel für ein solches Verfahren ist das *Interactive Multiple Goal Programming* (IMGP).

Bei den bisher aufgeführten Verfahrensklassen kann die Ausprägung eines Attributes durch entsprechende Ausprägungen der anderen Attribute kompensiert werden. Bei Verfahren mit nicht-kompensatorischem Präferenzfunktional geht das nicht. Dort werden die Alternativen zunächst nach dem wichtigsten Ziel geordnet. Sind sie dann noch nicht unterscheidbar, wird das zweitwichtigste Ziel als Entscheidungsmerkmal herangezogen usw. Beispiele für solche Verfahren sind die Lexikographische Ordnung und die Lexikographische Zielprogrammierung (engl.: *Goal Programming*).

4.2.2 Verfahren ohne Präferenzfunktional

Verfahren, die ein Präferenzfunktional nicht voraussetzen, benötigen keine schwache Ordnung der Alternativen, sondern lassen Intransitivitäten und/oder Unvergleichbarkeiten explizit zu. Im Vergleich zu den funktionsbasierten Verfahren bezeichnet man diese als Relationen basiert bzw. als „Outranking“-Verfahren. Diese eignen sich insbesondere bei vorliegen mehrkriterieller Konflikte, d. h. der Entscheider kann sich zwischen zwei Alternativen nicht entscheiden, weil beide Alternativen ein ausgewogenes Verhältnis an Vor- und Nachteilen beinhalten und somit keine besser da steht. In vielen Fällen reicht es in solchen Situationen, wenn man ohne eine vollständige Ordnung aller Alternativen nur die beste identifiziert oder eine Anordnung von Alternativenklassen erzielt. Entsprechende Verfahren sind z. B. ORESTE, *ELiminacion Et Choix Traduisant la REalite* (ELECTRE), *Preference Ranking Organization METHods for Enrichment Evaluations* (PROMETHEE) und *Data Envelopment Analysis* (DEA).

4.3 Nutzwertanalyse

Das in Kapitel 5 entwickelte Entscheidungsmodell basiert auf der klassischen NWA. Diese wird hier daher im Detail dargestellt (zur Begründung der Wahl dieses Verfahrens siehe Abschnitt 5.1).

Wie im vorangegangenen Abschnitt bereits angedeutet, ist die NWA streng genommen eine Klasse von Verfahren. Diese unterscheiden sich in ihrer speziellen Art der Ermittlung von Wertfunktionen (Höhenpräferenz) und Gewichten (Artenpräferenz), die jedoch alle holistisch und von einander unabhängig bestimmt werden.

Diese Gemeinsamkeit erlaubt es, von der Nutzwertanalyse zu sprechen (Schneeweß 1991, S. 121). Diese ist ein auf einem Präferenzindex basierendes Verfahren, mit dem Ziel ein Ranking der betrachteten Alternativen zu erzeugen.

Voraussetzungen für die Anwendung der NWA sind Vollständigkeit, Transitivität und schwache Präferenzunabhängigkeit der betrachteten Zielattribute (siehe Definitionen 1, 2, 5). Eine Einbettbarkeit der betrachteten Alternativen in ein Kontinuum ist nicht gefordert. Die NWA ist daher insbesondere dann sinnvoll, wenn eine endliche Anzahl an Alternativen vorliegt und nicht beliebig viele weitere Zwischenformen verfügbar sind. Außerdem muss der Entscheidungsträger keine Substitutionsraten angeben. Ein Vergleich beliebiger Wertdifferenzen ist somit ebenfalls nicht gefordert. Die betrachteten Attribute müssen lediglich mindestens ordinale Messniveau haben. Die Bewertung erfolgt in drei Schritten: (1) Bestimmung der Höhenpräferenz, (2) Bestimmung der Artenpräferenz und (3) Wertaggregation.

- 1. Bestimmung der Höhenpräferenz:** Für jedes Zielattribut c_j wird eine Wertfunktion $v_j(x_j(a))$ bestimmt, die alle in den Alternativen vorhandenen Ausprägungen in eine Besser-Schlechter-Reihenfolge bringt. Da mindestens ordinale Messniveau der Attribute ausreicht, ist die Wertfunktion lediglich bis auf (streng) monoton wachsende Transformationen festgelegt: „Ein ‘irgendwie gearteter Nutzen’, der in der ‘Person des Entscheidungsträgers und im betrachteten Attribut liegt’, wird jedem Attribut [...] zugeordnet“ (Schneeweß 1991, S. 123). Diese Funktion kann daher auch als *ordinale Nutzenfunktion* bezeichnet werden. Bei einer ordinalen Nutzenfunktion zeigt der Vergleich zweier Nutzenwerte nur an, welcher zu präferieren ist. Es ist jedoch keine konkrete Aussage zum Abstand zwischen diesen Werten möglich. Aussagen wie „Alternative a (ein Schüler mit der Note 3) ist doppelt so gut wie Alternative b (ein Schüler mit der Note 6)“ sind daher bei ordinal skalierten Attributen nicht möglich.

Die für ein Attribut verwendete Nutzenfunktion wird als *partielle Nutzenfunktion* bezeichnet und als $u_j(x_j(a))$ notiert. Bei endlich vielen Attributausprägungen und Alternativen hat man damit alle Nutzenwerte ermittelt und erhält eine Entscheidungsmatrix analog zu der in Abbildung 4.3. In der Regel sind die ermittelten Werte auf das Intervall $[0,1]$ normiert.

- 2. Bestimmung der Artenpräferenz:** Jedem Zielattribut wird seiner Bedeutung entsprechend ein Gewicht w_j ($\sum_j w_j = 1$, $w_j \geq 0$) zugeordnet.

Durch die Vergabe von Gewichten wird der Nutzen eines Attributs mit dem der übrigen Attribute vergleichbar. Die Gewichtung erfolgt holistisch, d. h. durch ganzheitliche Betrachtung der Wichtigkeit eines Attributs in Bezug auf sämtliche übrigen. Da diese entkoppelt von der Bestimmung der Artenpräferenz ist, haben konkrete Attributausprägungen keinen Einfluss darauf. Gewichtungsquotienten können somit auch nicht als Substitutionsraten interpretiert werden (Schneeweß 1991, S. 123 f.).

- 3. Wertaggregation:** Ziel der Wertaggregation ist ein Ranking der Alternativen bezüglich aller Zielattribute. Dazu wird eine Mehrzielpräferenzfunktion konstruiert, die analog zur Einzielpräferenzfunktion die Alternativen über die Äquivalenz $a' \succsim a'' \Leftrightarrow \Phi(a') \geq \Phi(a'')$ in eine Reihenfolge bringt. Ist zumindest schwache Präferenzunabhängigkeit gegeben, so lässt sich eine solche Präferenzfunktion als Funktion aller partiellen Nutzenfunktionen $u_j(x_j(a))$ einer Alternative darstellen:

$$U^{\text{NWA}}(a) = f(u_1(x_1(a)), u_2(x_2(a)), u_3(x_3(a)), \dots, u_m(x_m(a))) . \quad (4.3)$$

Die partiellen Nutzenfunktionen $u(\cdot)$ sind hierbei das Maß für die Höhenpräferenz, und die Präferenzfunktion $f(\cdot)$ ist das Maß für die Artenpräferenz. Höhen- und Artenpräferenz werden in Form eines linearen Präferenzindex vereinigt. Dieser wird auch als *utility score* bezeichnet:

$$U^{\text{NWA}}(a) = \sum_{j=1}^m w_j \cdot u_j(x_j(a)) . \quad (4.4)$$

Die beste Alternative a^* ist definiert durch $U^{\text{NWA}}(a^*) \geq U^{\text{NWA}}(a)$ mit $a^* \neq a$ für alle $a^*, a \in A$.

Die „Einfachheit“ der NWA hat auch Nachteile. Die holistische Bestimmung der Gewichte kann zu inkonsistenten Präferenzurteilen führen. Das Argument, die Gewichtung im Nachhinein zu korrigieren, hebt den Zweck der NWA aus. Gerade über die Bestimmung der Gewichte soll eine Anordnung der Alternativen erreicht werden und nicht umgekehrt. Da der Präferenzindex auf Basis der endlichen Menge an betrachteten Alternativen konstruiert wird, kann sich bei strikter Ordinalität der Attribute eine einmal ermittelte Rangordnung durch das Hinzukommen einer neuen Alternative ändern, obwohl die Präferenzstruktur des Entscheidungsträgers nach wie vor gleich geblieben ist. Bei der NWA besteht also die Gefahr der Instabilität. Ausgenommen hiervon sind Attribute die auf Ratingskalen (z. B. Schulnoten) basieren. Die vorgegebenen Skalenpunkte werden durch das Hinzuziehen weiterer Alternativen nicht verändert. Entsprechende Attribute können als „quasi-kardinal“ bezeichnet werden (Schneeweis 1991, S. 44). Sind alle Attribute quasi-kardinal, ist eine Einbettung weiterer Alternativen in den bestehenden Alternativenraum ohne Veränderung des bestehenden Rankings möglich.

4.4 Zusammenfassung

In diesem Kapitel wurde eine Einführung zu Ansätzen der Multikriteriellen Entscheidungsanalyse gegeben. Wichtige Verfahren wurden vorgestellt und die klassische Nutzwertanalyse wurde im Detail beschrieben.

MCDA ist für Entscheidungsprobleme vorgesehen, in denen zwischen verschiedenen Alternativen auf Grundlage mehrerer wichtiger Kriterien bzw. Attribute gewählt werden muss und bei denen diese Entscheidungsfindung durch eine einfache Beurteilung aus dem „Stegreif“ nicht möglich ist. Die zu erfüllenden Voraussetzungen für das Lösen des Entscheidungsproblems sind dabei abhängig vom eingesetzten Verfahren. Diese lassen sich in zwei Hauptkategorien teilen: Verfahren mit und ohne Präferenzfunktional. Bei Verfahren mit Präferenzfunktional wird vorausgesetzt, dass mindestens schwache Präferenzunabhängigkeit vorliegt und dass der Entscheidungsträger in der Lage ist, die Alternativen vollständig und transitiv anzurufen. Ziel ist es, ein Ranking der betrachteten Alternativen zu erhalten. Verfahren ohne Präferenzfunktional setzen diese Restriktionen nicht voraus. Bei diesen Verfahren wird an Stelle eines Rankings nur die Identifikation der besten bzw. der Menge der besseren Alternativen angestrebt.

Die NWA gehört zur Klasse der Verfahren mit Präferenzfunktional. Die Besonderheit der NWA liegt in der holistischen und von einander unabhängigen Ermittlung von Höhen- und Artenpräferenz. Die Attribute müssen mindestens ordinales Messniveau haben, Substitutionsraten müssen nicht angegeben werden.

Die NWA ist recht unkompliziert, hat aber auch Schwächen. Bei ordinal skalierten Attributen können nur Aussagen zu Besser-Schlechter-Verhältnissen ohne genaue Angaben von Abständen gemacht werden. Kommt zu einer endlichen Menge an Alternativen eine neue hinzu, kann sich das Rankingergebnis daher ändern, obwohl die Präferenzvorstellungen des Entscheidungsträgers gleich geblieben sind. Ausgenommen von dieser Instabilität sind kardinale und quasi-kardinale Attribute.

Kapitel 5

Bewertungsmodell für Webportale

Vorliegendes Kapitel beschreibt die Entwicklung eines multiattributiven Entscheidungsmodells zur Bewertung nicht-kommerzieller Informationsportale. Das Modell besteht aus drei Hauptbewertungskategorien: Nutzerverhalten, Nutzermeinungen und Aufwendungen.

Abschnitt 5.1 beinhaltet die zugrunde gelegte Bewertungsstrategie. In Abschnitt 5.2 wird in Bezug zur Strategie und für den hier betrachteten Kontext ein spezifischer Evaluationsansatz auf Basis des in Kapitel 3 entwickelten Rahmenmodells konzipiert. In Abschnitt 5.3 wird dem entsprechend das Entscheidungsmodell konstruiert. Abschnitt 5.4 fasst das Kapitel zusammen.

5.1 Bewertungsstrategie

Die zu wählende Bewertungsstrategie steht im engen Zusammenhang mit dem herangezogenen Bewertungsmaßstab. Für eine zielbasierte Bewertung sind tiefgehendes Wissen und Erfahrung zur Festlegung realistischer Ziele erforderlich. Im Fall nicht-kommerzieller Informationsportale ist das Vorhandensein entsprechenden Wissens nicht bekannt. Aufgrund der vorliegenden Literatur können ebenfalls keine bestehenden Standards identifiziert werden, die als Anhaltspunkte für eine normative Bewertung dienen könnten. Eine Bewertung mittels Vergleich mit eigenen Leistungen aus der Vergangenheit bildet keinen Bezug zur Außenwelt und ermöglicht daher keine realistische Einschätzung der Leistungsfähigkeit. Eine sinnvolle und zu aussagekräftigen Resultaten führende Bewertungsalternative im Fall nicht-kommerzieller Informationsportale wird hier nur in der komparativen Analyse gesehen, d. h. durch Vergleich mit anderen bestehenden Portalen. Diese erfolgt im Rahmen eines Benchmarkingprozesses mittels des Instrumentariums der NWA (siehe Abschnitt 4.3). Benchmarking kann wie folgt definiert werden:

Benchmarking ist die Suche nach den besten Industriepraktiken, die zu Spitzenleistungen führen. (Camp 1994, S. 16)

Benchmarking ist eine Form der komparativen Analyse, bei der die eigenen Leistungswerte mit denen anderer verglichen werden. So können realistische Rückschlüsse auf das eigene Leistungsniveau gezogen werden und bestehende Schwächen und Stärken identifiziert werden. Daneben können langfristig gute (engl.: *good practices*) bzw. beste (engl.: *best practices*) Vorgehensweisen identifiziert werden. Benchmarking ermöglicht somit nicht nur eine realistische Leistungseinschätzung, sondern auch die Aneignung von Know-how zur Verbesserung der eigenen Performance. Auf diese Weise lässt sich in einem kontinuierlichen Verbesserungsprozess langfristig ein hohes Leistungsniveau erzielen.

Benchmarking hat aber auch seine Grenzen. Voraussetzung dafür ist das Vorhandensein weiterer Organisationen und deren Bereitschaft zur Teilnahme am Benchmarkingprozess. Das wird insbesondere bei konkurrierenden Organisationen aus Angst vor möglichem Verlust eines Wissensvorsprungs eher schwer umzusetzen sein. Die reine Orientierung an Benchmarks kann außerdem nur dann zu besseren Praktiken führen, wenn diese in anderen Organisationen vorhanden sind. Ein allen vorhandenen Praktiken überlegenes Verfahren kann mit diesem Bewertungsansatz nicht identifiziert werden.

Der Benchmarkingprozess lässt sich in fünf Phasen unterteilen (Elmuti u. a. 1997):

1. **Planung:** Gegenstand der Planungsphase ist die Identifikation der strategischen Ziele des organisationalen Betriebs und der konkreten (Teil-)Prozesse, die analysiert werden sollen. Hilfreich wäre hierbei die Kenntnis von Nutzerbedürfnissen bezüglich dieser Prozesse. Darauf aufbauend ist ein Evaluationsansatz mit den entsprechenden Bewertungsperspektiven, -methoden und

-kriterien zu entwickeln. Daneben sind weitere Organisationen zu identifizieren, die als Benchmarks dienen können.

Eine detaillierte Beschreibung des entwickelten Evaluationsansatzes in Bezug auf nicht-kommerzielle Informationsportale erfolgt im anschließenden Abschnitt 5.2.

2. **Bildung eines Teams:** Das Benchmarking erfolgt in Zusammenarbeit mit Vertretern der teilnehmenden Organisationen. Die Minimalbesetzung besteht aus dem Evaluator selbst sowie jeweils mindestens einer Kontaktperson pro Portal.
3. **Datenerfassung:** In Zusammenarbeit mit den Vertretern der teilnehmenden Organisationen werden die benötigten Daten beschafft und für die Auswertung aufbereitet.
4. **Datenauswertung:** Die Datenauswertung erfolgt durch Benchmarking auf Basis der aus den verwendeten Kriterien ermittelten Performance-Kennzahlen. Als Benchmarkingmethode wird die klassische NWA eingesetzt. Diese erlaubt neben einer systematischen Vorgehensweise auch den Einbezug persönlicher Präferenzvorstellungen des Entscheidungsträgers in den Benchmarkingprozess. Stakeholderspezifische Interessen bezüglich der verwendeten Bewertungskriterien können auf diese Weise einbezogen werden.

Die NWA wird hier den anderen bestehenden MCDA-Verfahren aus mehreren Gründen vorgezogen:

- Ziel der Analyse ist ein Ranking der betrachteten Portale. Verfahren ohne Präferenzfunktional, wie z. B. ELECTRE, können daher nicht eingesetzt werden.
- Kardinales Messniveau liegt nicht bei allen hier verwendeten Bewertungskriterien vor. Angaben zu Substitutionsraten zwischen den verschiedenen Ausprägungen solcher Kriterien sind nicht möglich. Verfahren mit Präferenzfunktion, wie z. B. MAUT, können daher nicht eingesetzt werden.
- Die Menge der vorhandenen Alternativen ist begrenzt, d. h. es sind nicht beliebig viele Zwischenformen verfügbar. Verfahren, die einen unbegrenzten Alternativenraum voraussetzen, wie z. B. IMGP, können daher nicht eingesetzt werden.
- Verfahren, die einen Präferenzindex und einen begrenzten Alternativenraum voraussetzen, können hier angewendet werden. Der AHP ist eine mögliche Alternative. Diese Methode ermöglicht sogar die Bestimmung von zu konsistenten Situationen führenden Wertfunktionen, ist aber durch die notwendigen paarweisen Vergleiche sehr zeitaufwendig. Den Entscheidungsträgern in der hier vorliegenden Fallstudie (Kapitel 6) war dies aus Zeitgründen nicht möglich. Dementsprechend wurde als

Verfahren die klassische NWA gewählt. Bei dieser kann die Bestimmung der Gewichtungsfaktoren recht einfach vorgenommen werden.

Die Konstruktion des Entscheidungsmodells wird im Detail in Abschnitt 5.3 beschrieben.

5. **Konzeption von Handlungsansätzen:** Die Ergebnisse ermöglichen eine realistische Einschätzung des Erfolgs sowie die Identifikation vorhandener Schwachstellen eines Portals in Bezug auf die verwendeten Bewertungskriterien und einbezogenen Präferenzvorstellungen. Die gewonnenen Erkenntnisse helfen im Idealfall bei der Einschätzung von Handlungsbedarf zu bestimmten Schwachpunkten und – sofern good oder best practices identifiziert wurden – bei der Konzipierung von möglichen Verbesserungsansätzen.

5.2 Evaluationsansatz

Das Bewertungskonzept wird auf Basis des in Abschnitt 3.2 vorgestellten Rahmenmodells - der IS-Evaluationspyramide (Abbildung 3.1) - entwickelt.

5.2.1 Gegenstand und Aktivitätsbereich

Zunächst wird festgelegt, was hier unter „nicht-kommerziellen“ Webportalen zu verstehen ist:

Definition 6 (Nicht-kommerzielles Webportal). Ein nicht-kommerzielles Webportal bietet seine Inhalte und Dienstleistung allen interessierten Nutzern frei zugänglich und in der Regel ohne Kosten an. Die Generierung von Profit ist nicht Ziel des Betriebs.

Gegenstand der Bewertung sind nicht-kommerzielle Webportale mit Schwerpunkt in der kostenfreien Bereitstellung von Informationen. Dem vorgestellten Klassifikationsansatz entsprechend (Abbildung 2.2) sind das Portale mit dem Hauptzweck „Information“. Diese werden hier als „Informationsportale“ bezeichnet. Bei entsprechenden Portalen kommt es in erster Linie auf eine gute Qualität der bereitgestellten Informationen sowie der Informationsbereitstellung an. Daneben müssen sie aber auch gut erreich- und bedienbar sein, damit die Nutzer die gesuchten Informationen komfortabel abrufen können.

Die oben genannten Punkte können als wichtige Ziele solcher Portale angenommen werden. Der betrachtete Aktivitätsbereich ist somit die Bereitstellung und Pflege von Informationen sowie einer entsprechenden Portalinfrastruktur.

Da der Betrieb eines Portals erhebliche Kosten verursachen kann, ist eine Überprüfung der Zielerreichung in obigen Punkten sinnvoll und für mögliche Verbesserungen auch notwendig.

Für den Benchmarkingprozess sind weitere „ähnliche“ Portale heranzuziehen. Unter „ähnlichen“ Portalen sollen hier Portale verstanden werden, die größere Übereinstimmungen in der Themenwahl der bereitgestellten Inhalte, bei den Funktionen und Tools sowie bei den Zielgruppen haben.

5.2.2 Bewertungsperspektive

Zur Ermittlung der einzubehandelnden Bewertungsperspektiven lassen sich drei wichtige Stakeholdergruppen identifizieren: (1) Der Portalbetreiber, (2) die Portalnutzer sowie (3) der Geldgeber. Nach der Typologie von Mitchell u. a. (1997) (Abbildung 3.2) können diese drei Gruppen anhand der Attribute „Macht und Einfluss“, „Legitimität des Anspruchs“ und „Dringlichkeit des Anspruchs“ in folgende Stakeholdertypen untergliedert werden:

1. Portalbetreiber besitzen alle drei Attribute. Sie haben die Macht und den Einfluss, Portalinhalte und -funktionalität nach ihren Zielen und Wünschen zu gestalten. Als Betreiber ist dieser Machtanspruch auch legitim. Letztlich ist auch die Dringlichkeit gegeben, um die gesetzten Ziele durch den Portalbetrieb im vorgesehenen Zeitrahmen zu erreichen. Der Portalbetreiber hat insgesamt einen sehr starken Einfluss auf den Portalbetrieb und ist somit der Stakeholdertypologie entsprechend ein *maßgeblicher* (engl.: *definitive*) Stakeholder.
2. Portalnutzer, die zu einer der anvisierten Zielgruppen gehören, besitzen mindestens die Attribute „Legitimität“ und „Dringlichkeit“ des Anspruchs. Die Legitimität begründet sich darin, dass ein Nutzer einen gewissen Anspruch an die Qualität der gebotenen Dienste stellen kann, um diese Nutzen bringend verwenden zu können. Wird dieses Qualitätsniveau nicht erreicht, wird das Portal nicht im gewünschten Umfang genutzt und der Portalbetreiber hat seine Ziele verfehlt. Die Dringlichkeit ergibt sich aus dem Umstand, dass Informationen sowie Dienste nur dann Nutzen bringend sein können, wenn sie innerhalb des benötigten Zeitrahmens bereitgestellt werden. Informationen beispielsweise sind ab einem bestimmten Zeitpunkt veraltet und nicht mehr hilfreich.

Das Attribut „Macht und Einfluss“ wird ein einzelner Nutzer (unter vielen tausend Portalbesuchern) in der Regel nicht haben. Ausnahme wäre hier ein einer Gewerkschaft ähnlicher Zusammenschluss vieler Nutzer. Dieser Fall ist aber eher unwahrscheinlich und wird daher hier nicht betrachtet.

Nach der Stakeholdertypologie ergibt sich aus den zwei hier zugeordneten Attributen für einen Portalnutzer der Typ *abhängiger* (engl.: *dependent*) Stakeholder. Ein einzelner Nutzer hat somit keinen großen Einfluss. Dennoch ist ein Betreiber im Normalfall bestrebt, den Großteil seiner Klientel zufrieden zu stellen, damit ein erfolgreicher Portalbetrieb ermöglicht wird.

3. Der Gruppe der Geldgeber lassen sich mindestens die Attribute „Macht und Einfluss“ sowie „Legitimität“ zuordnen. Diese besitzen sie durch die Verfügungsgewalt über die finanziellen Ressourcen. Das Attribut „Dringlichkeit“ wird hier nicht zugeordnet, da der Geldgeber in der Regel nicht in konkrete (eventuell zeitkritische) Projekte des Portalanbieters eingebunden ist, sondern mit diesem eher langfristige, strategische Ziele vereinbart. Der Stakeholdertypologie entsprechend, ist der Geldgeber als *dominanter* (engl.: *dominant*) Stakeholder zu klassifizieren. In seinem Einflussbereich liegt die Entscheidungsgewalt darüber, ob der Portalbetrieb fortgeführt bzw. weiterhin finanziert wird oder nicht und welchen strategischen Zielen dieser unterliegt. Auf die Ausprägung konkreter Projekte in Zusammenhang mit dem Betrieb wird der Geldgeber in der Regel weniger Einfluss nehmen.

Von den drei hier als wichtig identifizierten Stakeholdergruppen werden der Portalbetreiber und die Portalnutzer in die Analyse einbezogen. Eine Betrachtung der Geldgeber ist hier nicht vorgesehen, ausgehend von der Annahme, dass zwischen Geldgeber und Portalbetreiber eine gewisse Abstimmung vorliegt und der Betreiber somit bereits im Auftrag und im Sinne des Geldgebers agiert.

5.2.3 Bewertungszweck

Folgende Ziele unterliegen dem in dieser Arbeit entwickelten Evaluationsansatz: (1) das Erfolgsniveau eines Portals realistisch einschätzen, (2) vorhandene Schwachstellen identifizieren sowie (3) Möglichkeiten zu Verbesserungsansätzen erkennen. Entsprechend den in Abschnitt 3.2.3 eingeführten Begrifflichkeiten ist der Bewertungszweck demnach eine *formative* Evaluation.

5.2.4 Zeitrahmen

Der Bewertungsansatz ist sowohl im Rahmen einer einmaligen Untersuchung, als auch in einer Langzeitstudie denkbar. Die primären Zielsetzungen in dieser Arbeit (siehe Punkt „Bewertungszweck“) können mittels einer einmaligen Untersuchung erreicht werden. Die in dieser Arbeit vorgenommene Fallstudie erfolgt daher im Rahmen einer einmaligen Datenerfassung. Eine Langzeitstudie wäre eine mögliche Zukunftsoption, um die Auswirkungen vorgenommener Änderungen zu überprüfen und um weiterführende Verbesserungen zu konzipieren.

5.2.5 Bewertungsmaßstab

Benchmarking als Bewertungsstrategie entspricht dem Ansatz der komparativen Analyse. Die Ergebnisse aller in die Untersuchung einbezogenen Portale dienen als Bewertungsmaßstab bzw. Benchmarks für die Interpretation der Ergebnisse des jeweils betrachteten Portals.

5.2.6 Analyseebenen

Aus den einbezogenen Stakeholdergruppen ergeben sich zwei Ebenen auf denen Daten erfasst und analysiert werden können: Die individuelle Ebene eines Portalnutzers und die organisationale Ebene eines Portalanbieters. Aus der individuellen Perspektive eines Nutzers lassen sich persönliche Meinung und Eindrücke zum Portal gewinnen. Die Betrachtung aus der organisationalen Perspektive des Portalanbieters erlaubt die Erfassung von Bereitstellungsprozessen von Portalinhalten.

Eine weitere Ebene, die in Betracht gezogen wird, ist die Systemebene des Portals. Hier werden die objektiv erfassten Daten zum Nutzerverhalten analysiert.

Alle drei betrachteten Ebenen (Individuum, Organisation, Portalsystem) ergänzen sich zu einem umfassenderen Betrachtungsfeld, als es auf Basis nur einer Ebene realisierbar wäre. So besteht auch die Möglichkeit, Zusammenhänge zwischen verschiedenen Ebenen (z. B. Nutzermeinungen und Nutzerverhalten) zu erkennen und daraus hilfreiche Erkenntnisse für die Evaluation zu gewinnen.

5.2.7 Methoden der Datenerfassung und -analyse

Benchmarking funktioniert nur auf Basis quantitativer Daten. Es können daher nur Methoden herangezogen werden, die Daten entsprechenden Typs liefern.

Die Analyse erfolgt im Rahmen einer Fallstudie, mit der Konsequenz der in Abschnitt 3.2.7 beschriebenen Vor- und Nachteile. Wie dort ebenfalls beschrieben, ist die Fallstudie genau genommen keine eigenständige Methode der Datenerfassung und -analyse, sondern beinhaltet die Anwendung einer oder mehrerer entsprechender Methoden in Kombination.

Für jede betrachtete Analyseebene wird hier eine spezielle Methode eingesetzt. Die individuelle Perspektive des Nutzers wird mittels eines strukturierten Fragebogens mit vorgegebenen Antwortkategorien erfasst. Individuelle Eindrücke sind so bis zu einem gewissen Grad quantitativ messbar. Die organisationale Perspektive des Portalfreibers wird mittels eines strukturierten Interviews erfasst. Ziel ist es, möglichst präzise, quantitative Aussagen zu Aufwendungen für die Pflege von Inhalten zu erhalten. Die Analyse der Nutzung auf Portalsystemebene erfolgt mittels Auswertung von Webserver-Logdaten.

Die zur jeweiligen Methode herangezogenen Bewertungskriterien werden nun im folgenden Abschnitt detailliert beschrieben.

5.2.8 Bewertungskriterien

Aufbauend auf oben vorgenommenen Festlegungen zu Aktivitätsbereich, Analyseebenen und -methoden werden hier drei Bereiche zur Entwicklung von Kriterien betrachtet: (1) die Qualität der bereitgestellten Informationen, (2) die Qualität des Portalsystems in Bezug auf die Informationsbereitstellung sowie (3) die Aufwendungen zur Bereitstellung und Pflege von Inhalten.

Ausgangspunkt für die Definition von Bewertungskriterien für diese Bereiche ist die Fragestellung nach deren Relevanz für nicht-kommerzielle Informationsportale.

Die in Abschnitt 3.2.8 beschriebenen Modelle stellen drei verbreitete Ansätze zu als relevant erachteten Variablen von IS-Erfolg und deren gegenseitiger Abhängigkeiten dar (siehe Tabelle 3.1). Mit der Variablenauswahl wird implizit bestimmt, welche Analyseebenen ebenfalls als relevant zu erachten sind. Das D/M92 (Abbildung 3.6) ist hierbei das einzige, welches neben der individuellen Ebene auch die organisationale Ebene sowie die technische Systemebene einbezieht. Dieses Modell wird daher hier als Grundlage zur Definition von Bewertungskriterien herangezogen.

Im Folgenden werden die Kriterien entsprechend den drei fokussierten Analyseebenen spezifiziert und erläutert.

Erfassung von Nutzerverhalten

Die Systemnutzung ist eine der sechs im D/M92 als wichtig für die Erfassung von IS-Erfolg erachteten Kategorien (repräsentiert durch die Kategorie *use* in Abbildung 3.6).

Wie in Abschnitt 3.3.2 eingehend erklärt, ist es grundsätzlich möglich, Nutzerverhalten durch Befragung oder mittels Beobachtung (z.B. durch Auswertung von Webserver-Logdaten) zu analysieren. Da letzteres in der Regel genauere und detailliertere Ergebnisse liefert, wird dieser Ansatz hier angewendet. Auf Basis der aus den Logdaten extrahierten Informationen werden die Klickpfade der Portalbesucher rekonstruiert und ausgewertet.

Eine sinnvolle Analyse von Nutzerklickpfaden (*web usage mining*) ist nur in Zusammenhang mit einer Analyse von Inhalten (*web content mining*) und Strukturen (*web structure mining*) möglich (siehe Abschnitt 3.3.2). Daher werden hier alle drei Aspekte des Web Mining bei der Entwicklung von Bewertungskriterien berücksichtigt.

Größere Websites und insbesondere Portale können viele tausend Seiten beinhalten. Das bedeutet, dass einzelne konkrete Seiten unter Umständen nur äußerst selten abgerufen werden. Eine Untersuchung der Klickpfade auf Ebene konkreter Zugriffe wird daher in der Regel zu keinen signifikanten bzw. brauchbaren Aussagen führen. Um entsprechende Informationen zu erhalten, müssen die konkreten Zugriffe abstrahiert werden. Dies geschieht hier durch *Generalisierung* (siehe Tabelle 3.7) auf Basis von Konzepthierarchien (siehe Abschnitt 2.2), die sich auf die Inhalte der abgerufenen Seiten beziehen und so eine Klassifikation ermöglichen. Im Anschluss daran werden die generalisierten Klickpfade analysiert.

Konzepthierarchien In der Regel ist eine Kombination verschiedener Konzepthierarchien notwendig, um die interessanten Inhalte und Zusammenhänge einer Website widerzuspiegeln. Die im Modell vorgeschlagenen Hierarchien bilden die technischen Services des Portals, den Typ einer angefragten Portalseite sowie den Typ der Referrerseite, von der aus die Anfrage kam, ab. Abbildung 5.1 beinhaltet die hier definierten Konzepthierarchien und deren Beziehungen zu einander. Die Buchstabenkombinationen in Klammern enthalten die Kurzbezeichnungen für

das jeweilige Konzept. Im Weiteren werden diese Kurzbezeichnungen verwendet. Detaillierte Beschreibungen zum jeweiligen Typ von Service (Dienst), Seite und Referrer finden sich in den Tabellen 5.1, 5.2 und 5.3.

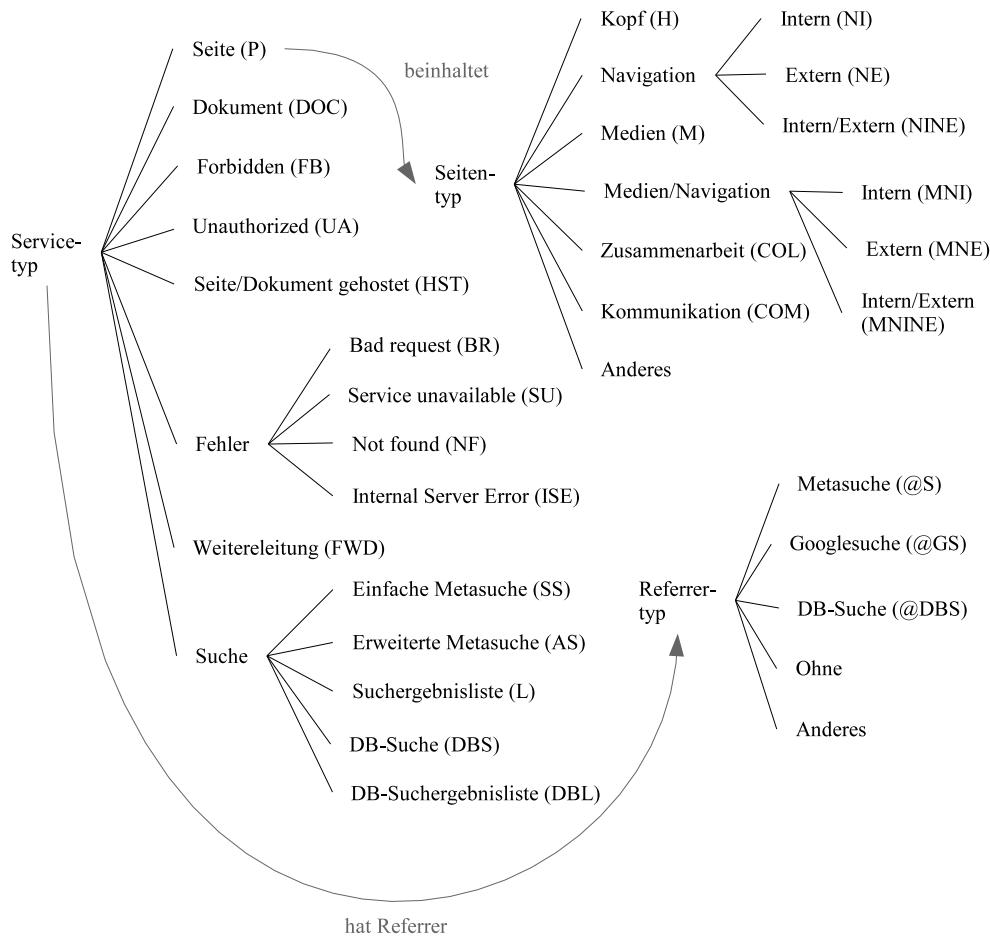


Abbildung 5.1: Konzepthierarchien zur Analyse des Nutzerverhaltens

Wenn ein Nutzer bzw. Portalbesucher einen Request mittels Webbrowser an das Portal sendet (z.B. durch Klicken auf einen Link), erfolgt eine Serviceanfrage an das Portal. Dieses „antwortet“ dann abhängig von der Art der Anfrage entweder mit einer normalen Seite oder einem herunterladbaren Dokument oder einer Suchergebnisliste oder einer Fehlermeldung usw. Was auf einen spezifischen Request zurückgesendet wurde, lässt sich im Log am URI-Eintrag, Referrereintrag und am HTML-Status-Code identifizieren (siehe dazu die Tabellen 3.6, 5.1, 5.2 und 5.3). Normale Seiten, die zurückgesendet werden, können bezüglich Inhalt und Funktion weiterhin in verschiedene Typen klassifiziert werden. Die Klassifikation erfolgt durch Auswertung des jeweiligen URI- und Referrereintrags im Log. Wenn die URI dafür nicht genügend Informationen enthält, wird das Hyperlink-zu-Text-

Verhältnis der entsprechenden Portalseite ausgewertet und damit der Seitentyp bestimmt.

Tabelle 5.1: Betrachtete Servicetypen

Typ	Service	Beschreibung
P	Seite	Seite eines bestimmten Typs.
DOC	Dokument	Herunterladbares Dokument (z. B. eine PDF-Datei).
HST	Gehostet	Seite oder herunterladbares Dokument gehostet oder eine diesbezügliche Fehlermeldung.
BR	Info: Bad Request	Html-Fehlermeldung: Falsche Anfrage (Status-Code 400).
UA	Info: Unauthorized	Html-Fehlermeldung: Authentifizierung notwendig (Status-Code 401).
FB	Info: Forbidden	Html-Fehlermeldung: Zugang nicht erlaubt (Status-Code 403).
NF	Info: Not Found	Html-Fehlermeldung: Seite/Dokument nicht gefunden (Status-Code 404).
ISE	Info: Internal Server Error	Html-Fehlermeldung: Serverfehler (Status-Code 500).
SU	Info: Service Unavailable	Html-Fehlermeldung: Server temporär nicht erreichbar (Status-Code 503).
SS	Einfache Suche	Einfaches Suchformular.
AS	Erweiterte Suche	Suchformular mit erweiterten Suchoptionen.
L	Suchergebnisliste	Liste mit den Suchergebnissen zur einfachen oder erweiterten Suche.
DBS	DB-Suche	Datenbank-Suchformular.
DBL	DB-Suchergebnis	Liste mit den Suchergebnissen zur DB-Suche.
FWD	Weiterleitung	Funktion zur automatischen Weiterleitung einer Anfrage von einer Seite zu einer anderen.

Klickpfadanalyse Die Klickpfadanalyse erfolgt auf Basis der klassifizierten Requests, d. h. es werden die generalisierten Pfade betrachtet. Wie im Weiteren noch gezeigt wird, beinhaltet die Auswertung der Pfade implizit auch eine Auswertung der Portalstruktur. Damit sind alle drei obigen Aspekte des Web Mining bei der Analyse des Nutzerverhaltens berücksichtigt.

Gegenstand der hier durchgeführten Klickpfadanalyse ist eine Untersuchung, inwieweit der portalinterne Menüaufbau sowie die Verlinkungsstruktur der Seiten eine gute Navigation innerhalb des Portals ermöglichen und an welchen Stellen Verbesserungen vorgenommen werden könnten (Ziel des Einsatzes von Web Usage Mining ist hier somit die Websiteanpassung; siehe Abbildung 3.11). Betrachtet wird

Tabelle 5.2: Betrachtete Seitentypen

Typ	Seite	Beschreibung
H	Kopf-/Startseite	Startseite (z. B. Homepage) des betrachteten Portalbereichs.
NI, NE, NINE	Navigations-Seite	NI: Enthält Links zu ausschließlich internen Zielen. NE: Enthält Links zu ausschließlich externen Zielen. NINE: Enthält Links zu internen und externen Zielen. Charakterisiert durch ein großes Link-zu-Text-Verhältnis.
M	Medien-Seite	Enthält Informationscontent. Charakterisiert durch ein kleines Link-zu-Text-Verhältnis.
MNI, MNE, MNINE	Medien/Navigations-Seite	Enthält sowohl Informationscontent als auch analog zu den reinen Navigations-Seiten weiterführende Links. Charakterisiert durch ein mittleres Link-zu-Text-Verhältnis.
COM	Kommunikations-Seite	Beinhaltet eine Funktionalität zur Unterstützung der Kommunikation zwischen Nutzern (z. B. ein Forum).
COL	Zusammenarbeits-Seite	Beinhaltet eine Funktionalität zur Unterstützung der Zusammenarbeit zwischen Nutzern (z. B. BSCW).

hierbei die Episode (siehe Tabelle 3.5) bzw. der Teil der Klickpfade zwischen dem ersten Zugriff auf die Homepage (Typ H) und dem darauf folgenden ersten Zugriff auf eine Zielseite aus der Menge $\mathcal{P} = \{M, MNI, MNE, MNINE\}$ der betrachteten Seitentypen bzw. auf ein herunterladbares Dokument (Typ DOC).

Genaugenommen können sowohl Seiten als auch herunterladbare Dokumente als „Webserverdokumente“ bezeichnet werden. Im Folgenden wird daher mit den Begriffen „Webserverdokument“ bzw. „Zieldokument“ ein Zugriff aus der Menge $\mathcal{Z} = \{\mathcal{P}, \text{DOC}\}$ bezeichnet. Von Interesse ist, wie die Klickpfade derjenigen Nutzer aussehen, die für entsprechende Zugriffe ausschließlich das vorgegebene Menü und die auf den Seiten vorhandenen Hyperlinks nutzen und welche Schlüsse aus diesen Pfaden gezogen werden können. Dazu wird eine Sequenzanalyse der im Log

Tabelle 5.3: Betrachtete Referrertypen

Typ	Referrer	Beschreibung
@S	Interne Metasuche	Aktuelle Anfrage basiert auf Suchergebnis der portalinternen Meta-Suchmaschine.
@GS	Googlesuche	Aktuelle Anfrage basiert auf Suchergebnis einer Googleabfrage.
@DBS	Interne DB-Suche	Aktuelle Anfrage basiert auf Suchergebnis einer portalinternen DB-Suchmaschine.

abgespeicherten und generalisierten Nutzungsmuster durchgeführt (siehe Berendt u. Spiliopoulou 2000; Spiliopoulou u. Pohle 2001).

Ein betrachteter Satz an Logdaten enthält eine Menge $\mathcal{S} = \{s_1, s_2, \dots, s_N\}$ an Besucher- bzw. Nutzersessions. Jede Session besteht wiederum aus einer Menge $s = \{r_1, r_2, \dots, r_L\}$ an Requests.

Definition 7 (Aktive Session). Eine Session $s \in \mathcal{S}$ mit mindestens einem Request r_i , welcher dem Konzept H zugeordnet werden kann (dargestellt mit $con(r_i) = H$), ist eine aktive Session.

Da nur die von der Homepage ausgehenden Klickpfade untersucht werden, sind nur jene Sessions interessant, die mindestens einen Homepage-Request beinhalten, d.h. nur aktive Sessions werden betrachtet. Die Menge aller aktiven Sessions ist entsprechend Definition 7 bestimmt durch

$$\mathcal{S}^{\text{ACT}} = \{s \in \mathcal{S} | con(r_i) = H \wedge r_i \in s\} \quad (5.1)$$

für alle $i = 1, \dots, L$.

Definition 8 (Subsequenz). Sei $seq(s) = \langle a_1 a_2 \dots a_n \rangle$ mit $n \in \mathbb{N}$ eine Sequenz, d.h. eine vollständig geordnete Liste, aller Requests in Session s . Dann ist $sseq(s) = \langle b_1 b_2 \dots b_m \rangle$ eine Subsequenz von $seq(s)$, dargestellt mit $sseq(s) \sqsubset seq(s)$, genau dann wenn ein i existiert mit $a_i, \dots, a_{(i+m-1)} \in s$ und $a_{(i+j-1)} = b_j$ für alle $j = 1, \dots, m$.

Die interessierende Subsequenz eines Klickpfades in einer Besuchersession beginnt mit dem ersten Zugriff auf eine Seite des Typs H und endet mit dem ersten Zugriff auf ein Zieldokument des Typs $z \in \mathcal{Z}$. Eine entsprechende Subsequenz sei als z -Subsequenz bezeichnet und dargestellt mit $z-sseq(s)$. Formal ausgedrückt ist $z-sseq(s)$ eine Subsequenz $sseq$ der Session s , so dass

$$con(b_1) = con(a_i) = H \quad (5.2)$$

mit $i = \min_{j=1,2,\dots,n} \{j | con(a_j) = H\}$ und

$$con(b_m) = con(a_{i+m-1}) = z \quad (5.3)$$

mit $m = \min_{j=1,2,\dots,n-1} \{j | con(a_{i+j-1}) = z\}$.

Folgende Ausnahme wird für den Fall $z \neq \text{DOC}$ bei der Betrachtung von Subsequenzen gemacht:

Fragt ein Nutzer beispielsweise 10 herunterladbare Dokumente von einer und derselben Seite an, so hat er 10 Zugriffe erzeugt, ohne sich in dem Portal weiter fortbewegt zu haben. Aus diesen Zugriffen können also keine Rückschlüsse zur Menü- und Portalstruktur gezogen werden. Für den Fall $z \neq \text{DOC}$ werden daher in der betrachteten Subsequenz vorhandene DOC-Zugriffe nicht mitgezählt. Der

Wert von m wird dementsprechend um die Anzahl an DOC-Zugriffen in dieser Subsequenz verringert.

Ein Zugriff, der auf einer vorangegangenen Suchmaschinen- oder DB-Abfrage basiert, ist anhand von zwei Kennzeichen identifizierbar: (1) Es ist ein Request nach einer Suchergebnisliste ($con(r_i) = \{L, DBL\}$) und/oder (2) der zugehörige Referrereintrag zeigt an, dass der Request aus einer Suchanfrage resultiert. In diesem Fall kann dieser auf Grund des Referrereintrags, dargestellt mit $refcon(r_i)$, einem Konzept aus der Menge $\mathcal{R} = \{@S, @GS, @DBS\}$ zugeordnet werden.

Definition 9 (Positive Session). Eine Session $s \in \mathcal{S}$ mit der Subsequenz $z\text{-sseq}(s)$, in welcher keine Suchmaschinen- oder DB-Requests enthalten sind, ist eine positive Session bezogen auf den Typ z des abgerufenen Zieldokumentes.

Nutzer, die eine Suchmaschine (portalintern oder extern) oder eine Datenbank zum Auffinden von Seiten verwenden, sind nicht abhängig vom Navigationsmenü und der physikalischen Linkstruktur des Portals. Entsprechende Klickpfade können daher keinen Aufschluss zur Qualität von Menü- und Portalstruktur geben. Solche *nicht-positive* Sessions sind somit hier uninteressant und werden im Weiteren nicht betrachtet. Die Menge aller positiven Sessions mit Bezug auf z ist entsprechend Definition 9 bestimmt durch

$$\mathcal{S}_z^{POS} = \left\{ s \in \mathcal{S} \mid z\text{-sseq}(s) \Vdash seq(s) \wedge con(r_i) \neq \{L, DBL\} \wedge refcon(r_i) \neq \{@S, @GS, @DBS\} \right\} \quad (5.4)$$

für alle $r_i \Vdash z\text{-sseq}(s)$.

Definition 10 (Effektivität von z -Requests). Die Effektivität von Requests nach einem speziellen Zieldokumententyp z relativ zu allen aktiven Sessions ist bestimmt durch

$$eff(z) = \frac{|\mathcal{S}_z^{POS}|}{|\mathcal{S}^{ACT}|}. \quad (5.5)$$

Der Effektivitätswert gibt an, wie groß der Anteil eines bestimmten Zieldokumententyps in den aktiven Sessions ist. Je höher ein solcher Wert ist, desto mehr Sessions enthalten entsprechende Zugriffe.

Definition 11 (Effizienz von z -Requests). Die Länge einer z -Subsequenz in $s \in \mathcal{S}$, dargestellt als $length(z\text{-sseq}(s))$, wird bestimmt durch die Anzahl der darin enthaltenen Elemente, H nicht mitgezählt: $length(z\text{-sseq}(s)) = m_s - 1$. Die durchschnittliche Effizienz von Requests nach einem Zieldokument vom Typ z , bezogen auf alle positiven Sessions mit einem entsprechenden Request, ist dann bestimmt durch

$$efz(z) = \frac{|\mathcal{S}_z^{POS}|}{\sum_{s \in \mathcal{S}_z^{POS}} length(z\text{-sseq}(s))}. \quad (5.6)$$

Der Effizienzwert ist ein Indikator dafür, wie viele Seiten durchschnittlich in den positiven Sessions angefragt werden müssen, bevor der erste Zugriff auf ein Zieldokument des Typs z erfolgt. Je höher dieser Wert ist, desto kürzer sind die entsprechenden Klickpfade im Durchschnitt.

Proposition 1 *Der Effektivitäts- und der Effizienzwert für Zugriffe auf ein Zieldokument eines bestimmten Typs sind Indikatoren für das entgegengebrachte Interesse seitens der Nutzer sowie für Schwachstellen in der Menü- und Portalstruktur.*

Ziel- bzw. Webserverdokumente mit hohem Effektivitäts- und hohem Effizienzwert werden oft angefragt, sind leicht zu finden und auf relativ kurzen Pfaden von der Homepage aus zu erreichen (z.B. mittels eines Links auf der Homepage). Man kann hier von der Annahme ausgehen, dass entsprechende Dokumente interessant für die Nutzer sind oder aber zu anderen interessanten Dokumenten führen. Hohe Effektivitäts- und Effizienzwerte weisen außerdem darauf hin, dass entsprechende Webserverdokumente im Portal gut positioniert sind. Dokumente mit hohem Effektivitäts- jedoch niedrigem Effizienzwert werden zwar oft angefragt, sind aber umständlich zu erreichen (z. B. durch eine komplizierte Menüführung). Entsprechend zu obiger Annahme kann man hier davon ausgehen, dass deren Positionierung im Portal oder aber die Menüführung an dieser Stelle verbesserungsfähig ist. Bei Dokumenten mit niedriger Effektivität, unabhängig vom Effizienzwert, ist dann analog zu obiger Annahme davon auszugehen, dass die Inhalte uninteressant sind oder aber dass diese einfach nur schwer zu finden sind (z. B. komplizierte und verwirrende Menüstruktur). Aus dem Nutzerverhalten alleine lässt sich dies nicht eindeutig beantworten.

Die gezogenen Schlüsse sind zur Übersicht in der Matrix in Abbildung 5.2 nochmals zusammengefasst.

	hohe Effektivität	niedrige Effektivität
hohe Effizienz	<ul style="list-style-type: none"> • leicht zu finden • auf kurzen Pfaden erreichbar • Content ist gefragt 	<ul style="list-style-type: none"> • auf kurzen Pfaden erreichbar • Content ist uninteressant ODER schwer zu finden
niedrige Effizienz	<ul style="list-style-type: none"> • umständlich erreichbar • Content ist gefragt 	<ul style="list-style-type: none"> • umständlich erreichbar • Content ist uninteressant ODER schwer zu finden

Abbildung 5.2: Effektivitäts-Effizienz-Matrix

Aus der Analyse des Nutzerverhaltens lassen sich keine Informationen über die Zufriedenheit der Nutzer zu den Inhalten der angefragten Zieldokumente ableiten. Dafür ist eine direkte Befragung notwendig.

Erfassung von Nutzermeinungen

Im vorliegenden Abschnitt wird die Entwicklung des Fragebogens zur Erfassung der Nutzermeinungen beschrieben.

Basis für die Herleitung der relevanten Kriterien zur Gestaltung der Fragen ist das D/M92 in einer vereinfachten Form (Abbildung 5.3): Die *use*-Komponente wird nicht einbezogen, da diese sich auf das Nutzerverhalten und nicht auf Nutzermeinungen bezieht; des Weiteren werden organisationale Auswirkungen (*organizational impact*) im vorliegenden Ansatz nicht betrachtet.

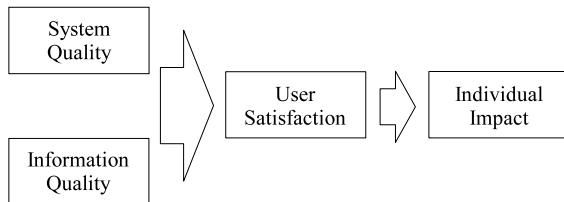


Abbildung 5.3: Vereinfachtes D/M92

Der direkteste Weg, den Erfolg eines Portals auf Basis der Nutzermeinungen zu bewerten, geht über eine Erfassung der individuell wahrgenommenen Auswirkungen (*individual impact*). In den meisten Fällen sind diese aber besonders schwer zu messen. Einfacher zu analysieren sind dagegen die Konstrukte „System- bzw. Websitequalität“ (*system quality*), „Informationsqualität“ (*information quality*) oder „Nutzerzufriedenheit“ (*user satisfaction*) (siehe Abschnitt 3.2.8). Obwohl Messergebnisse auf Basis dieser Konstrukte meist eine gute Approximation der Auswirkungen darstellen, kann man diese nicht pauschal und uneingeschränkt als Ersatz heranziehen, da die Abhängigkeiten zwischen diesen nicht immer ausreichend stark sind (Yuthas u. Young 1998). Wenn möglich sollten daher immer die Auswirkungen direkt erfasst werden. Wie DeLone u. McLean (2003) aber anmerken, können Auswirkungen nicht vollständig verstanden und interpretiert werden, ohne eine Betrachtung von Ursachen. Daher ist es wichtig, möglichst viele Konstrukte bzw. IS-Erfolgskategorien in den Analyseprozess mit einzubeziehen. Der Fragebogen wird somit aus Fragen zusammengesetzt, die sich auf alle im vereinfachten D/M92 enthaltenen Kategorien beziehen.

Zur Erfassung der Kategorien „Systemqualität“ (*system quality*), „Informationsqualität“ (*information quality*) und „Nutzerzufriedenheit“ (*user satisfaction*) werden verbreitet angewendete Kriterien aus bestehenden und bereits getesteten Ansätzen übernommen (siehe Aladwani u. Palvia 2002; Liu u. Arnett 2000; Negash u. a. 2003). Aus der Schwierigkeit heraus, Auswirkungen direkt zu erfassen bzw. Kriterien zur Erfassung von Auswirkungen zu definieren (siehe Abschnitt 3.2.8), beziehen die meisten hier bekannten Ansätze sie nicht ein. In einigen Studien lassen sich aber mindestens vier Faktoren identifizieren, die durch Wahrnehmung von Systemqualität, Informationsqualität oder Nutzerzufriedenheit beeinflusst werden können und daher als potentielle Kriterien zur Messung individueller Auswirkungen

in Frage kommen. Diese sind (1) Vertrauen, (2) empfundenes Risiko, (3) Loyalität und (4) Wissen. Detaillierte Erläuterungen zu diesen Faktoren wurden bereits in Abschnitt 3.2.8 gemacht. Sie werden hier noch einmal im Kontext einer Portalanalyse beschrieben:

1. **Vertrauen:** Damit ein Portal überhaupt genutzt wird, muss der Nutzer Vertrauen in das technische System und dessen Betreiber haben. Nur so wird er bereit sein, den Informationsinhalten Glauben zu schenken, persönliche Daten zu übermitteln und die angebotenen Dienste zu nutzen. Bei Informationsportalen ist es besonders wichtig, dass der Portalbetreiber als glaubwürdig und kompetent bezüglich der Inhalte angesehen wird. Vertrauen ist somit Voraussetzung für einen erfolgreichen Portalbetrieb.

Nutzer, die mit einem Portal und dessen Betreiber noch nicht sehr vertraut sind, bilden ihr Vertrauen auf der Wahrnehmung bestimmter vertrauensbildender Attribute. Verschiedene Arbeiten zeigen, dass hierfür insbesondere Qualitätsmerkmale der Website ausschlaggebend sind, wie z. B. die Informationsqualität, die Funktionalität des technischen Systems oder die Ästhetik des Oberflächendesigns (z. B. Bart u. a. 2005; Corritore u. a. 2005; Fogg u. a. 2002; Gefen 2002a; Kim 2000; McKnight u. a. 2002; Warrington u. a. 2000).

2. **Risiko:** Ein Portal zu nutzen, birgt grundsätzliche Risiken, wie z. B. Desinformation oder den Verlust der persönlichen Privatsphäre. Je nachdem, wie stark das Risikoempfinden beim Nutzer ist, wird dieser eine entsprechende Menge an Vertrauen benötigen, um sein Risikoempfinden zu kompensieren und das Portal zu nutzen (z. B. Gefen 2002a; Mayer u. a. 1995).

3. **Loyalität:** Loyalität gilt als einer der wichtigsten Erfolgsfaktoren im eCommerce (Reicheld u. Shefter 2000). Loyale Kunden sind langfristig profitabler, produzieren mehr und regelmäßig Umsatz und verbreiten positive Publicity für den Verkäufer. Analog dazu lassen sich loyale Mitglieder einer nicht-kommerziellen Portalcommunity definieren: Loyale Nutzer besuchen regelmäßig das Portal, tragen durch eigene Beiträge, z. B. in einem Forum, zum erfolgreichen Portalbetrieb bei und sind bereit, das Portal anderen weiter zu empfehlen.

Untersuchungen zeigen, dass Loyalität durch wahrgenommene Qualität der Website sowie durch Vertrauensbildung aus wahrgenommener Qualität beeinflusst wird (z. B. Flavián u. a. 2006; Gefen 2002a; Warrington u. a. 2000).

4. **Wissensbildung:** Ein weiteres wichtiges Kriterium für die Erfassung von Auswirkungen bei Informationsportalen ist der Einfluss der konsumierten Informationen auf das Wissen des Nutzers (siehe D'Ambra u. Rice 2001). In Abhängigkeit von der Qualität der bereitgestellten Informationen wird ein Nutzer dementsprechend sein Wissen mehr oder weniger stark erweitern können.

Die in den vier obigen Punkten angeführten Arbeiten bestätigen Beziehungen zwischen Qualitätsmerkmalen einer Website auf der einen Seite und der Bildung von Vertrauen, Risikoempfinden, Loyalitätsgefühl sowie Wissen auf der anderen. Diese Ergebnisse lassen folgende Annahme zu:

Proposition 2 „Vertrauen“, „empfundenes Risiko“, „Loyalität“ und „Wissen“ sind Konstrukte, die durch empfundene Systemqualität und Informationsqualität beeinflusst werden. Diese Konstrukte können daher als individuelle Auswirkungen der Portalnutzung interpretiert werden.

Die im D/M92 postulierten Beziehungen zwischen der Kategorie „Individuelle Auswirkungen“ und den restlichen im Modell aufgeführten Kategorien lassen das in Abbildung 5.4 dargestellte Pfadmodell für die Erfolgsbewertung von Portalen annehmen. Dieses beinhaltet die Operationalisierung der Kategorie „Individuelle Auswirkungen“ entsprechend Proposition 2 und stellt die angenommenen Abhängigkeiten zwischen diesen Operationalisierungskonstrukten auf Grundlage der Ergebnisse der oben in den vier Punkten aufgeführten Arbeiten dar. Die angenommenen Abhängigkeiten zwischen der Kategorie „Individuelle Auswirkungen“ und den restlichen im Pfadmodell aufgeführten Kategorien basieren ebenfalls zum Teil auf diesen sowie auf Untersuchungsergebnissen weiterer Arbeiten (siehe DeLone u. McLean 2003).

Aus dem Pfadmodell ergeben sich die in Tabelle 5.4 dargestellten Hypothesen. Grundsätzlich ist zu erwarten, dass sich System- und Informationsqualität positiv auf die Zufriedenheit und die individuell wahrgenommenen Effekte auswirken. Hohe System- und Informationsqualität führt zu steigender Zufriedenheit (H1, H2), zu steigendem Vertrauen dem Betreiber gegenüber (H4, H7, H10) sowie zu steigender Wissensbildung (H3, H6, H9). Der Einfluss auf das empfundene Risiko verhält sich dementsprechend umgekehrt (H5, H8, H11). Hohes Vertrauen und gestiegene Wissensbildung führen zu steigender Loyalität gegenüber dem Portal (H13, H14). Zunehmendes Vertrauen mindert gleichzeitig das empfundene Risiko (H12). Hohes Risikoempfinden mindert die empfundene Loyalität (H15).

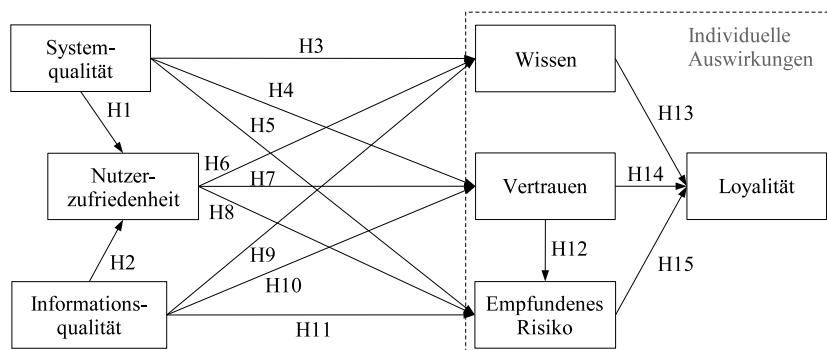


Abbildung 5.4: Pfadmodell zur Erfassung der Nutzermeinungen

Tabelle 5.4: Hypothesen zu Konstruktabhängigkeiten

Nr.	Hypothese
H1	Es besteht ein positiver Zusammenhang zwischen empfundener Systemqualität und Nutzerzufriedenheit.
H2	Es besteht ein positiver Zusammenhang zwischen empfundener Informationsqualität und Nutzerzufriedenheit.
H3	Es besteht ein positiver Zusammenhang zwischen empfundener Systemqualität und Wissensbildung.
H4	Es besteht ein positiver Zusammenhang zwischen empfundener Systemqualität und Vertrauensbildung.
H5	Es besteht ein negativer Zusammenhang zwischen empfundener Systemqualität und dem empfundenen Risiko.
H6	Es besteht ein positiver Zusammenhang zwischen Nutzerzufriedenheit und Wissensbildung.
H7	Es besteht ein positiver Zusammenhang zwischen Nutzerzufriedenheit und Vertrauensbildung.
H8	Es besteht ein negativer Zusammenhang zwischen Nutzerzufriedenheit und dem empfundenen Risiko.
H9	Es besteht ein positiver Zusammenhang zwischen empfundener Informationsqualität und Wissensbildung.
H10	Es besteht ein positiver Zusammenhang zwischen empfundener Informationsqualität und Vertrauensbildung.
H11	Es besteht ein negativer Zusammenhang zwischen empfundener Informationsqualität und empfundenem Risiko.
H12	Es besteht ein negativer Zusammenhang zwischen Vertrauensbildung und empfundenem Risiko.
H13	Es besteht ein positiver Zusammenhang zwischen Wissensbildung und Loyalität.
H14	Es besteht ein positiver Zusammenhang zwischen Vertrauensbildung und Loyalität.
H15	Es besteht ein negativer Zusammenhang zwischen empfundenem Risiko und Loyalität.

Werden die Annahmen des Modells im Rahmen einer empirischen Analyse weitgehend bestätigt, kann dies als ein positives Indiz für die hier gewählten Kriterien zur Erfassung von Auswirkungen gewertet werden.

Für den Fragebogen wurden die zu den Kriterien entsprechenden Fragen aus den oben genannten Ansätzen übernommen. Die Fragen wurden aus dem englischen übersetzt und unverändert oder bei Bedarf in adaptierter Form verwendet. Für die endgültige Formulierung wurden vier Studenten, zwei Graduierte und zwei Professoren gebeten, diese auf Verständlichkeit zu prüfen und ggf. Korrekturen vorzuschlagen. Die endgültige Version des Fragebogens (Tabelle 5.5) besteht aus insgesamt 21 Fragen zu den Kategorien

- Informationsqualität (INFQ1–INFQ5),
- Systemqualität (SYSQ1–SYSQ6),

- Zufriedenheit (ZUFR1, ZUFR2) und
- Individuelle Auswirkungen (IAUS), operationalisiert durch
 - Wissen (WISS1, WISS2),
 - Vertrauen (VERT1, VERT2),
 - Risiko (RISK1, RISK2) und
 - Loyalität (LOYA1, LOYA2).

Die Antworten werden auf einer unipolaren Likert-Ratingskala (Likert 1932) mit fünf Kategorien von „1“ („stimme überhaupt nicht zu“) bis „5“ („stimme voll zu“) erfasst. Ziel der unipolaren Skalierung ist es, den Grad des wahrgenommenen Effektes zur jeweiligen Frage zu ermitteln. Dieser kann von nicht wahrgenommen („stimme überhaupt nicht zu“) bis sehr stark wahrgenommen („stimme voll zu“) gehen. Eine Skalierung mit fünf Stufen ist in vielen Ansätzen üblich und wurde hier gewählt, um die Differenzierungsfähigkeit der urteilenden Personen nicht zu überlasten.

Mit Ratingskalen kann man in der Regel sehr schnell und unkompliziert Urteile erfassen und auswerten. Der Einsatz von Ratingskalen hat aber auch Nachteile. Obwohl die Abstände zwischen den Stufen nicht definiert sind, wird oft von gleichen Abständen ausgegangen, d. h. eine Ratingskala wird oft als intervallbasierte Skala interpretiert (Bortz u. Döring 2006, S. 176 ff.). Es bleibt allerdings offen, ob alle Umfrageteilnehmer intervallskalierte Urteile abgeben. Es besteht außerdem die Gefahr, dass die Umfrageteilnehmer kein einheitliches Verständnis zu Einheit und Ursprung der genutzten Skala haben. Das kann dazu führen, dass Objekte (Fragen) mit extremer Merkmalsausprägung von einigen Teilnehmern nicht mehr korrekt eingestuft werden können, weil die Extremwerte zuvor bereits für andere Objekte (Fragen) mit weniger extremen Merkmalsausprägungen vergeben wurden (Ceiling- oder Flooreffekt) (Bortz u. Döring 2006, S. 182). Es können auch systematische Urteilsfehler auftreten: Der Urteilende versäumt, konzeptuell unterschiedliche und potenziell unabhängige Merkmale zu differenzieren (Haloeffekte); die Beurteilung fällt systematisch zu positiv bzw. zu negativ aus (Milde-Härte-Fehler); der Urteilende vermeidet prinzipiell extreme Beurteilungen (Tendenz zur Mitte) usw. (Bortz u. Döring 2006, S. 183 ff.). Grundsätzlich lassen sich solche Fehler nicht komplett vermeiden. Durch Aufklärung und eine gute Einführung der Befragten in die Thematik der Umfrage lassen sie sich zumindest minimieren.

Der Fragebogen enthält zusätzlich noch einen demographischen Teil. Erfragt werden Nutzergruppenzugehörigkeit, Erfahrung mit dem Internet sowie mit dem Portal, Nutzungshäufigkeit des Portals und Interessenschwerpunkte (Tabelle 5.6).

Erfassung von Aufwendungen

Um dem Veralten von Inhalten vorzubeugen, ist regelmäßige Contentpflege notwendig. Je mehr Contentpflege betrieben wird, desto aktueller sind die Informationen, desto mehr Aufwand ist dafür aber auch notwendig.

Tabelle 5.5: Fragebogen

Informationsqualität (INFQ)	
INFQ1	Die Informationen sind aktuell.
INFQ2	Die Informationen sind verständlich.
INFQ3	Die Informationen sind vollständig.
INFQ4	Die Informationen sind für mich relevant.
INFQ5	Die Informationen sind exklusiv.
Systemqualität (SYSQ)	
SYSQ1	Der Portalthalt ist gut strukturiert und organisiert.
SYSQ2	Der Portalthalt ist gut lesbar.
SYSQ3	In diesem Portal lässt es sich einfach navigieren.
SYSQ4	Die Links (Hyperlinks) funktionieren.
SYSQ5	Die einzelnen Seiten (Pages) werden schnell aufgebaut.
SYSQ6	Die Übermittlung von sensiblen Daten (z.B. Passwörter) ist sicher.
Zufriedenheit (ZUFR)	
ZUFR1	Insgesamt bin ich zufrieden mit diesem Portal.
ZUFR2	Insgesamt erfüllt dieses Portal meine Erwartungen.
Individuelle Auswirkungen (IAUS)	
	Wissen (WISS)
WISS1	Ich bin besser informiert.
WISS2	Die Informationen helfen mir beim Treffen von Entscheidungen.
	Vertrauen (VERT)
VERT1	Ich vertraue xyz.de, dass sie ihren Job richtig machen.
VERT2	Dieses Portal ist in jeder Beziehung vertrauenswürdig.
	Risiko (RISK)
RISK1	Es ist unsicher, mit xyz.de zu interagieren.
RISK2	Es besteht ein deutliches Risiko in der Nutzung dieses Portals.
	Loyalität (LOYA)
LOYA1	Ich beabsichtige, dieses Portal in Zukunft regelmäßig zu besuchen.
LOYA2	Ich werde xyz.de weiter empfehlen.

Mit der Auswertung von Nutzerverhalten und Nutzermeinungen werden Effekte des Portalbetriebs betrachtet. Die Erfassung der Aufwendungen als eine weitere Komponente soll Hinweise dazu geben, wie effizient der Portalbetrieb in Bezug auf die erzielten Effekte gestaltet wird. Dahinter steht die Überlegung, dass die Pflege von Inhalten am effizientesten ist, wenn die Frequenz von Änderungen der Inhalte der Besuchsfrequenz wiederkehrender Nutzer entspricht. Wenn ein Nutzer also einmal pro Woche das Portal besucht, ist es vergebene Mühe, für diesen zweimal pro Woche Content zu aktualisieren. Die Hälfte der Änderungen wird er nicht mitbekommen. Auf der anderen Seite würde er schnell gelangweilt, wenn der Content nur alle zwei Wochen geändert wird. Die mögliche Konsequenz wäre der Verlust dieses Nutzers.

Eine derart genaue Analyse war in der Fallstudie auf Basis des vorliegenden

Tabelle 5.6: Demographischer Teil des Fragebogens

Ich bin...	
1	SchülerIn
2	ReferendarIn
3	LehrerIn
4	DirektorIn
5	Elternteil
6	MitarbeiterIn einer außerschulischen Bildungseinrichtung
7	MitarbeiterIn der Bildungsverwaltung
8	keiner dieser Gruppen zugehörig
Ich nutze des World Wide Web seit...	
1	weniger als einem Jahr
2	1–2 Jahren
3	3–5 Jahren
4	mehr als 5 Jahren
Ich nutze dieses Portal seit...	
1	weniger als einem Monat
2	1–3 Monaten
3	4–12 Monaten
4	mehr als 12 Monaten
Ich besuche dieses Portal...	
1	fast täglich
2	mindestens einmal pro Woche
3	mehrmals monatlich
4	gelegentlich
Welche Punkte motivieren Sie besonders, das Portal wiederholt zu Besuchen?	
1	Informationen zu bildungspolitischen Ereignissen
2	Informationen zu Schulen und Schulfächern
3	Informationen zu Bildungs- und Fortbildungsangeboten
4	Download von Informationsangeboten (z. B. Rahmenlehrpläne, Unterrichtsmaterialien)
5	Kommunikation mit anderen Nutzern (z. B. Forum, Chat)
6	Kooperation mit anderen Nutzern (z. B. BSCW)
7	eLearning
8	Andere Gründe

Datenmaterials nicht durchführbar (siehe Kapitel 6). Dazu wäre ein Mechanismus nötig gewesen, der wiederkehrende Nutzer erkennt und diese Information in der Webserver-Logdatei festhält. Bei keinem der in der Fallstudie betrachteten Portale war ein solcher Mechanismus vorhanden. Es wird hier daher eine abgeschwächte Form der Aufwandsanalyse vorgenommen. Diese soll zusammen mit dem Ergebnis aus der Umfrage zumindest einen ersten Hinweis darauf liefern, ob die Aktualisierungshäufigkeit der Inhalte ausreichend ist. Außerdem sollte das Ergebnis einen ersten Eindruck vom durchschnittlich notwendigen Aufwand zur Portalpflege.

ge geben. Für die Auswertung werden die Häufigkeiten mit denen Inhalte erstellt, gelöscht und aktualisiert werden, betrachtet.

Definition 12 (Aufwendungen). *Die Aufwendungen des Portalbetreibers zur Pflege von Portalinhalten werden erfasst als die Anzahl der Vorgänge zur Erstellung neuen Contents (create), zur Änderung bestehenden Contents (update) und zur Löschung veralteten Contents (delete).*

Diese Zahlen können aber nicht einfach als Benchmark-Kriterien übernommen werden. Große Portale haben sehr viele Seiten und benötigen daher im Normalfall auch mehr Änderungsvorgänge bzw. Pflegeaufwand als kleine Portale. Ein simpler Vergleich dieser Zahlen würde also zum zwingenden Ergebnis führen, dass große Portale bei der Effizienzbetrachtung schlechter abschneiden. Um hier einen realistischen Vergleich durchführen zu können, werden diese Werte relativiert, indem man sie im Verhältnis zur Anzahl der Besuchersessions aus einem bestimmten Zeitraum betrachtet.

Definition 13 (Vorgangstyp). *Sei $V = \{v_1, v_2, v_3, \dots, v_z\}$ die Menge aller Vorgänge für einen betrachteten Zeitraum und V_{vtyp} die Menge aller Vorgänge eines bestimmten Typs $vtyp$ für diesen Zeitraum, mit $V_{vtyp} \subseteq V$ und $vtyp(v_i) = vtyp(v_j)$ für alle $v_{i,j} \in V_{vtyp}$, $i \neq j$ und $vtyp \in \{\text{create}, \text{update}, \text{delete}\}$.*

Das Verhältnis der Anzahl an Vorgängen eines bestimmten Typs zur Anzahl der Besuchersessions aus einem bestimmten Zeitraum kann als Indikator dafür gesehen werden, wie lohnend Änderungsvorgänge sind.

Definition 14 (Contentänderungsverhältnis). *Das Contentänderungsverhältnis cav für einen bestimmten Vorgangstyp $vtyp$ und der Menge \mathcal{S} aller betrachteten Sessions aus einem bestimmten Zeitraum ist gegeben durch*

$$cav(vtyp) = \frac{|V_{vtyp}|}{|\mathcal{S}|}. \quad (5.7)$$

Proposition 3 *Der Portalpflegeaufwand pro Besuchersession, gemessen als Contentänderungsverhältnis, ist ein Indikator dafür, wie bezahlt sich Vorgänge für Bereitstellung (create), Aktualisierung (update) und Löschung (delete) von Portalcontent in Bezug auf die Anzahl der Nutzer machen.*

Je mehr Besuchersessions ein Portal vorweist, desto besser ist das Contentänderungsverhältnis bezogen auf einen bestimmten Vorgang, d. h. desto geringer ist der Portalpflegeaufwand je realisierter Session. Dieses lässt sich als Effizienzkriterium des Portalbetriebs interpretieren, wenn man davon ausgeht, dass eine Änderung umso lohnender ist, je mehr Nutzer diese wahrnehmen. Dahinter steht eine ähnliche Überlegung wie bei der Vermarktung von Werbezeitraum in z. B. TV-Sendungen. Je mehr Zuschauer eine Sendung hat, desto mehr potentielle Kunden erreicht eine

Werbeeinblendung und desto teurer wird entsprechender Werbezeitraum verkauft. Der Einsatz einer Werbeeinblendung wird demnach als umso lohnender gesehen, je mehr Zuschauer diese potentiell verfolgen können.

Ein zu kleiner Wert des Contentänderungsverhältnisses für einen bestimmten Vorgang kann allerdings ein Indiz dafür sein, dass an der Stelle zu wenig Portalpflegeaufwand betrieben wird und die Gefahr besteht, dass Nutzer vermehrt auf veralteten Content stoßen.

5.3 Entscheidungsmodell

Der vorangegangene Abschnitt beschreibt detailliert den Aufbau des hier entwickelten Evaluationsansatzes für nicht-kommerzielle Informationsportale. Die definierten Bewertungskriterien werden im Entscheidungsmodell übernommen. Die Konstruktion des Modells beginnt mit der Aufstellung des Zielsystems. Das Oberziel ist die Erfassung von Webportalerfolg. Die nachfolgenden Operationalisierungsstufen basieren auf den drei im vorangegangenen Abschnitt definierten Bewertungskategorien „Nutzerverhalten“, „Nutzermeinungen“ und „Aufwendungen“ mit den dort definierten Kriterien. Es unterliegt die Annahme, dass diese (Ziel-) Kriterien die Bedingungen für Operationalität, (Ziel-)Vollständigkeit und Redundanzfreiheit erfüllen und dass keine systematischen Beziehungen zwischen diesen vorliegen (siehe Abschnitt 4.1). Die entsprechende Zielhierarchie ist in Abbildung 5.5 dargestellt.

5.3.1 Nutzerverhalten

Zielkriterien für die Komponente „Nutzerverhalten“ sind die Effektivitäts- ($eff(z)$) und Effizienzwerte ($efz(z)$) für die von der Homepage ausgehenden Zugriffe auf Zieldokumente des Typs $z \in \mathcal{Z}$ (siehe Konzepthierarchie in Abbildung 5.1 sowie die Tabellen 5.1 und 5.2). Die resultierende Entscheidungsmatrix beinhaltet somit zehn Kriterien (Tabelle 6.9): Einen Effektivitäts- und einen Effizienzwert je betrachtetem Zieldokumententyp. Die Anwendung einer Nutzenfunktion (siehe Abschnitt 4.1) zur Normalisierung der Ergebniswerte auf das Intervall $[0, 1]$ ist hier nicht nötig, da die Effektivitäts- und Effizienzwerte bereits normalisiert sind. Sie können direkt als partielle Nutzenwerte übernommen werden.

Den verwendeten Zielkriterien entsprechend wird der Utility Score für die Komponente „Nutzerverhalten“ bestimmt durch

$$U_{\text{Nutzerverhalten}}^{\text{NWA}}(a_i) = \sum_{j=1}^5 w_j \cdot \left[\sum_{k=1}^2 w_{jk} \cdot x_{jk}(a_i) \right], \quad (5.8)$$

für alle $a_i \in A$ und $x_{jk} \in \{eff(z), efz(z)\}$ mit j als Index für den Typ des Zieldokumentes und k als Index für den jeweils entsprechenden Effektivitäts- bzw. Effizienzwert (siehe Abbildung 4.3 und Tabelle 6.9).

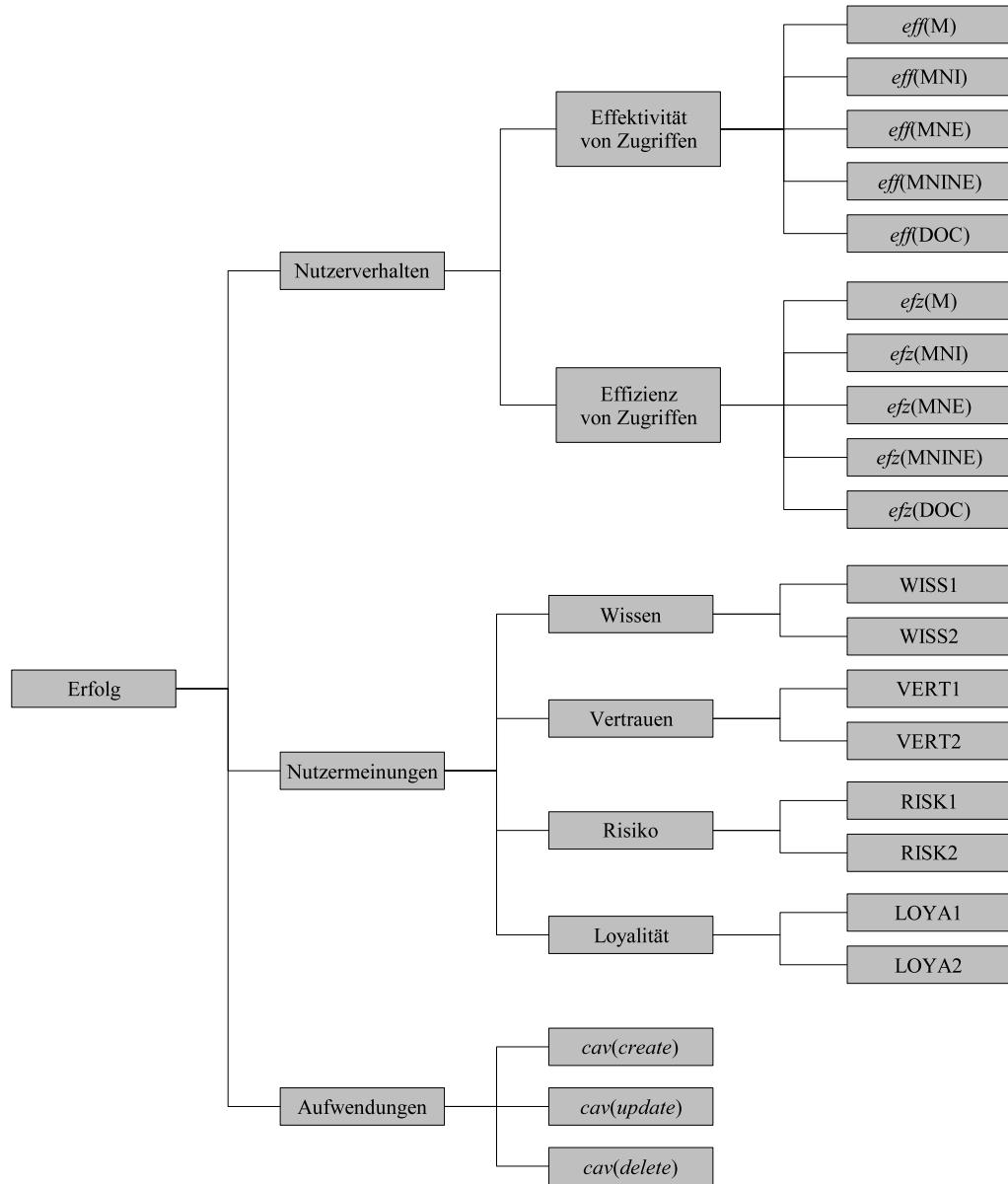


Abbildung 5.5: Zielhierarchie für Webportalerfolg

5.3.2 Nutzermeinungen

Als Zielkriterien für die Komponente „Nutzermeinungen“ werden die acht Fragen der Kategorie IAUS in Tabelle 5.5 zur Erfassung der individuellen Auswirkungen der Portalnutzung herangezogen (Delić u. Lenz 2008a). Diese operationalisieren die vier verwendeten Bewertungskonstrukte „Wissen“, „Vertrauen“, „empfundenes Risiko“ und „Loyalität“. Wie in Abschnitt 5.2.8 verdeutlicht, unterliegt hier die

Annahme, dass diese Fragen (bzw. Kriterien in der Entscheidungsmatrix) in ausreichendem Maße individuell wahrgenommene Auswirkungen erfassen.

Die Antworten werden mittels einer Ratingskala gemessen. Wie oben beschrieben, wird bei Ratingskalen oft die Annahme getroffen, dass diese als intervallbasierte Skalen interpretiert werden können. Da die Gefahr besteht, dass nicht alle Umfrageteilnehmer intervallskalierte Urteile abgeben, wird dieser Annahme hier nicht gefolgt und die Skala wird als Ordinalskala betrachtet. Als Konsequenz kann für die Herleitung der partiellen Nutzenwerte auf Basis der Antworten zu einer Frage nicht der Durchschnittswert der Ratings genommen werden. Stattdessen werden die Quartile herangezogen. So wird auch der Einfluss von Ausreißern auf das Ergebnis verringert. Ein Beispiel für einen Ausschnitt einer entsprechenden Ergebnismatrix ist in Tabelle 5.7 gegeben.

Tabelle 5.7: Beispieldausschnitt der quartilbasierten Ergebnismatrix zur Komponente „Nutzermeinungen“

c_j	WISS					
c_{jk}	WISS1			WISS2		
c_{jkl}	1. Quartil	Median	3. Quartil	1. Quartil	Median	3. Quartil
a_i	3	4	5	3	3	4

Zur Bestimmung der partiellen Nutzenwerte wird die in diesem Fall ordinale Nutzenfunktion

$$u(x_{jkl}(a_i)) = \frac{x_{jkl}(a_i)}{5} \quad (5.9)$$

mit $x_{jkl} \in \{1, 2, \dots, 5\}$ entsprechend den fünf Stufen der im Fragebogen verwendeten Ratingskala eingesetzt. Zu beachten ist hierbei, dass die abgeleiteten Nutzenwerte somit nur als ordinal skalierte Werte betrachtet werden dürfen. Ein Beispiel für einen Ausschnitt einer entsprechenden Entscheidungsmatrix ist in Tabelle 5.8 gegeben.

Tabelle 5.8: Beispieldausschnitt der quartilbasierten Entscheidungsmatrix zur Komponente „Nutzermeinungen“

w_j	0,242					
c_j	WISS					
w_{jk}	0,625			0,375		
c_{jk}	WISS1			WISS2		
w_{jkl}	0,33	0,33	0,33	0,33	0,33	0,33
c_{jkl}	1. Quartil	Median	3. Quartil	1. Quartil	Median	3. Quartil
a_i	0,6	0,8	0,1	0,6	0,6	0,8

Den verwendeten Zielkriterien entsprechend, wird der Utility Score für die Komponente „Nutzermeinungen“ bestimmt durch

$$U_{\text{Nutzermeinungen}}^{\text{NWA}}(a_i) = \sum_{j=1}^4 w_j \cdot \left[\sum_{k=1}^2 w_{jk} \cdot \left[\sum_{l=1}^3 w_{jkl} \cdot u(x_{jkl}(a_i)) \right] \right], \quad (5.10)$$

für alle $a_i \in A$ mit j als Index für die jeweilige Fragebogenkategorie, k als Index für die jeweilige Frage einer Fragebogenkategorie sowie l als Index für den jeweiligen quartilbasierten Nutzenwert zu einer Frage (siehe Tabelle 5.8).

5.3.3 Aufwendungen

Als Zielkriterien für die Komponente „Aufwendungen“ werden die Contentänderungsverhältnisse für Erstellen ($cav(create)$), Aktualisieren ($cav(update)$) und Lösen ($cav(delete)$) herangezogen. Die Nutzenfunktion zur Herleitung der partiellen Nutzenwerte ist

$$u(x_j(a_i)) = 1 - x_j(a_i) * 100 \quad (5.11)$$

mit $x_j \in \{cav(create), cav(update), cav(delete)\}$ und 100 als Skalierungsfaktor, damit die Werte nicht zu klein werden. Die resultierenden Ergebnisse werden wie folgt interpretiert: Je geringer ein Nutzenwert ist, desto größer ist der entsprechende Aufwand bzw. die Anzahl der Vorgänge ($create$, $update$ oder $delete$) im Verhältnis zur Anzahl der Besuchersessions.

Den verwendeten Zielkriterien entsprechend, wird der Utility Score für die Komponente „Aufwendungen“ bestimmt durch

$$U_{\text{Aufwendungen}}^{\text{NWA}}(a_i) = \sum_{j=1}^3 w_j \cdot u(x_j(a_i)), \quad (5.12)$$

für alle $a_i \in A$ mit j als Index für den jeweiligen Contentänderungstyp (siehe Tabelle 6.22).

5.3.4 Gesamtbewertung

Für die Berechnung des Gesamtergebnisses werden die Utility Scores der einzelnen Komponenten als partielle Nutzenwerte herangezogen, gewichtet und zu einem Gesamt-Utility Score summiert.

Definition 15 (Gesamt-Utility Score). Der Gesamt-Utility Score $U_{\text{Gesamt}}^{\text{NWA}}(a_i)$ zu einem Portal errechnet sich aus den Utility Scores der betrachteten Komponenten „Nutzerverhalten“, „Nutzermeinungen“ und „Aufwendungen“ durch

$$\begin{aligned} U_{\text{Gesamt}}^{\text{NWA}}(a_i) &= w_1 \cdot U_{\text{Nutzerverhalten}}^{\text{NWA}}(a_i) \\ &\quad + w_2 \cdot U_{\text{Nutzermeinungen}}^{\text{NWA}}(a_i) \\ &\quad + w_3 \cdot U_{\text{Aufwendungen}}^{\text{NWA}}(a_i), \end{aligned} \quad (5.13)$$

mit $\sum_{j=1}^3 w_j = 1$ und $w_j \geq 0$ entsprechend den Präferenzvorstellungen des Entscheiders.

Der Gesamt-Utility Score stellt einen Gesamtnutzenwert dar, der sich aus den drei hier betrachteten Bewertungskomponenten ableitet.

Definition 16 (Erfolg). Der Erfolg bzw. das Erfolgsniveau eines Informationsportals in der Bereitstellung von Inhalten, bezogen auf die betrachteten Kriterien und Präferenzvorstellungen des Entscheiders, ist bestimmt durch dessen Gesamt-Utility Score $U_{\text{Gesamt}}^{\text{NWA}}$ relativ zu den Ergebnissen der anderen Portale.

Aus den Definitionen 15 und 16 folgt, dass das erfolgreichste Portal, dargestellt mit a^* , den höchsten Gesamt-Utility Score besitzt:

$$U_{\text{Gesamt}}^{\text{NWA}}(a^*) \geq U_{\text{Gesamt}}^{\text{NWA}}(a_i), \quad (5.14)$$

mit $a^* \neq a_i$ für alle $a^*, a_i \in A$. Die Gewichtung spiegelt die Präferenzvorstellungen des Entscheiders bezüglich dieser drei Komponenten wider. Entsprechend diesen Vorstellungen und dem Bewertungsergebnis können Schlüsse gezogen werden, ob und in welchem Bereich eine Verbesserung vorgenommen werden sollte.

5.4 Zusammenfassung

In diesem Kapitel wurde ein spezifischer Evaluationsansatz und darauf aufbauend ein multiattributives Entscheidungsmodell zur Erfolgsbewertung nicht-kommerzieller Informationsportale entwickelt. Als grundlegende Bewertungsstrategie wurde Benchmarking gewählt, durchgeführt mittels der Nutzwertanalyse. Diese erlaubt einen systematischen und konsistenten Vergleich, basierend auf den Präferenzvorstellungen des Entscheidungsträgers.

Der dem Entscheidungsmodell zugrunde liegende Evaluationsansatz ist speziell für nicht-kommerzielle Informationsportale konzipiert. Das Konzept legt fest, welche Analyseebenen, -methoden und -kriterien zur Bewertung in Bezug auf die betrachteten Stakeholdergruppen herangezogen werden sollen. Drei Bereiche werden analysiert: das Nutzerverhalten, die Nutzermeinungen und die Aufwendungen zur Pflege der Inhalte. Die für diese Bereiche konzipierten Kriterien zur Datenerfassung werden als Zielkriterien im Entscheidungsmodell übernommen. Dieses besteht dementsprechend aus den Komponenten „Nutzerverhalten“, „Nutzermeinungen“ und „Aufwendungen“.

Kapitel 6

Fallstudie Bildungsserver

Vorliegendes Kapitel beschreibt die Anwendung des in Kapitel 5 entwickelten Bewertungsmodells im Rahmen einer Fallstudie von vier deutschen Bildungsserver-Portalen.

Abschnitt 6.1 gibt eine kurze Einführung zu den betrachteten Portalen. Abschnitt 6.2 beschreibt die Untersuchung zum Nutzerverhalten. Die Prozesse der Datenbeschaffung, Datenaufbereitung und Ergebnisherleitung werden dargestellt, gefolgt von der Auswertung, Interpretation und Diskussion der Ergebnisse. Die Abschnitte 6.3 (Nutzermeinungen) und 6.4 (Aufwendungen) sind analog aufgebaut. In Abschnitt 6.5 werden die Ergebnisse zu Nutzerverhalten, Nutzermeinungen und Aufwendungen in einem Gesamtmodell integriert und eine Gesamterfolgsbewertung wird vorgenommen. Abschnitt 6.6 beinhaltet die Schlussfolgerungen zu Verbesserungen, die sich aus der Untersuchung für einen der Portalbetreiber ergeben. Abschnitt 6.7 fasst das Kapitel zusammen.

6.1 Betrachtete Portale

Im Rahmen der Fallstudie wurden vier deutsche Bildungsserver-Portale untersucht. In Deutschland gibt es insgesamt 16 Landes- sowie einen gesamtdeutschen Bildungsserver. Ein Schwerpunkt des Betriebs dieser Portale ist die Bereitstellung von Informationen. Die Inhalte beziehen sich landesspezifisch auf bildungsrelevante Themen wie Bildung, Bildungspolitik, Schulen, Lehre, Lehrerfort- und -weiterbildung u. ä. Daneben werden verschiedene Dienste wie eLearning und Unterstützung für Onlinezusammenarbeit bereitgestellt. Die Gestaltung der Portale konzentriert sich auf die Inhalte, weniger auf Design und spezielle Effekte. Zu den Zielgruppen gehören Lehrer, andere Lehrende, Schüler, Eltern, Mitarbeiter außerschulischer Bildungseinrichtungen, Mitarbeiter der Bildungsverwaltung und andere an Bildung Interessierte.

Die Bildungsserver eignen sich besonders gut für eine Untersuchung im Rahmen eines Benchmarkingansatzes, da mehrere vorhanden sind und sie nicht in Konkurrenz zueinander stehen. Der Informationsaustausch im Rahmen einer Untersuchung ist daher unproblematisch. Daraus ergeben sich gute Chancen, dass teilnehmende Portale von der Identifikation von Schwachstellen und good oder best practices profitieren.

Um eine ausreichende Anzahl an Fallstudienobjekten zu erhalten, wurden 10 Betreiber von Bildungsservern kontaktiert. Das Anschreiben beinhaltete den Zweck der Untersuchung und einen Hinweis auf mögliche Nutzen bringende Ergebnisse für den weiteren Portalbetrieb. Um Teilnahme wurde gebeten, ohne weitere Anreize zu bieten. Vier Bildungsserver sagten einer Kooperation zu. Ein weiteres Portal hätte erst zu einem späteren Zeitpunkt teilnehmen können und konnte daher nicht mitberücksichtigt werden.

Einer Vereinbarung mit den Betreibern entsprechend, wird die Identität der teilgenommenen Portale nicht preisgegeben. Diese werden im Folgenden mit P1, P2, P3 und P4 bezeichnet. Um zumindest ein Bild vom betrachteten Portaltyp zu geben, wird hier als Beispiel der Schweizerische Bildungsserver (www.educa.ch) angeführt. Dieser ist den untersuchten Portalen in der Art der Gestaltung der Inhalte und Menüstrukturen ähnlich (siehe Abbildung 6.1).

Die Betreiber bzw. stellvertretend die mit dem Evaluator kooperierenden Kontaktpersonen der Portale werden hier als Entscheidungsträger für die Herleitung von Präferenzstrukturen herangezogen. Da die Präferenzvorstellungen eines Entscheiders in Form einer spezifischen Gewichtung der Kriterien in das Entscheidungsmodell einbezogen werden, sind die ermittelten Nutzenwerte subjektiv beeinflusst. Wenn Portalbetreiber P1 eine von P2 abweichende Präferenzstruktur hat, werden dementsprechend die Ergebnisse für diese beiden Portale unterschiedlich ausfallen. Um den Fallstudienbericht in einem akzeptablen Rahmen zu halten, wird die Bewertung hier nur aus der Perspektive des Betreibers von P2 vorgenommen. Für eine Auswertung aus den Perspektiven der anderen Provider müssten jeweils deren Gewichtungen der Kriterien gesetzt und die Ergebnisse erneut berechnet werden.

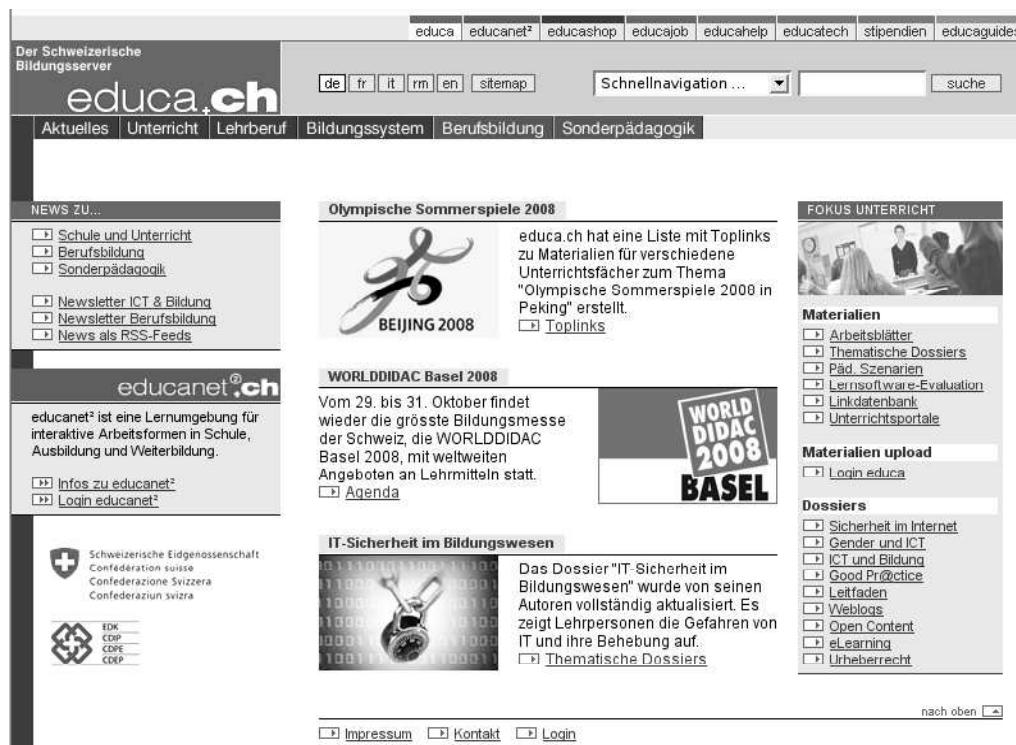


Abbildung 6.1: Homepage des Schweizerischen Bildungsservers vom 14.05.2008

In den folgenden Abschnitten werden die empirischen Resultate anhand des im vorangegangenen Kapitel entwickelten Entscheidungsmodells (siehe Abschnitt 5.3) erfasst und ausgewertet.

6.2 Bewertung des Nutzerverhaltens

Die Analyse des Nutzerverhaltens erfolgt auf Basis der auf den Webservern der Portale gespeicherten Logdaten. In der Untersuchung werden die Logdaten vom 15. und 19. November 2006 (P1, P2, P3) sowie vom 4. und 7. März 2007 (P4) analysiert.¹

Wie im vorangegangenen Kapitel 5 erläutert, werden nur Sessions mit einem Homepage-Request also nur aktive Sessions betrachtet. Diese werden der Einfachheit halber im Weiteren als „H-Sessions“ bezeichnet.

Ein Vergleich der durchschnittlichen Zugriffszahlen pro H-Session aus den zwei gewählten Tagen mit H-Sessions aus einem Zeitraum von vier Wochen weist keine großen Unterschiede auf (Tabelle 6.1). Man kann daher davon ausgehen, dass die ausgewählte Teilmenge an Sessions ausreichend groß ist, um hinreichend repräsentativ zu sein.

¹ Die Logdaten aus einem größeren Zeitraum zu analysieren, z. B. vier Wochen, war hier aufgrund des hohen Auswertungsaufwands nicht umsetzbar.

tative Ergebnisse zu liefern.

Tabelle 6.1: H-Sessions

Portal	H-Sessions (vier Wochen)	Zugriffe Durchschnitt (vier Wochen)	H-Sessions (zwei Tage)	Zugriffe Durchschnitt (zwei Tage)
P1	12297	7,1	746	6,9
P2	30888	7,9	2168	7,3
P3	67682	5,5	4692	5,4
P4 ^a	6186	10,5	1661	9,9

^aBei P4 standen Logdaten aus dem Zeitraum von nur einer Woche zur Verfügung.

6.2.1 Datenaufbereitung

In der Aufbereitungsphase werden alle Einträge aus dem Log entfernt, die für die Analyse nicht notwendig sind bzw. die Ergebnisse verfälschen könnten. Dazu gehören Zugriffe von Webrobots und Zugriffe auf unwichtige Dateien. Zur Webroboterkennung wird das Agent-Feld (siehe Tabelle 3.6) auf gängige Muster von Webrobotnamen (z. B. „bot“, „crawl“, „check“) untersucht. Zusätzlich wird ein Abgleich mit der im Web (www.robotstxt.org) frei verfügbaren Liste aller bekannter Webrobots durchgeführt. Es wird hier davon ausgegangen, dass die meisten Robot-zugriffe so erkannt werden. Weitere unwichtige Einträge sind Zugriffe auf Grafiken (z. B. *.bmp, *.jpeg, *.gif) sowie Konfigurationsdateien, Scripte u. ä. (z. B. *.css, *.js). Übrig bleiben nur die Einträge mit Requests nach Html-Seiten und herunterladbaren Dokumenten (z. B. *.pdf, *.doc).

Sessionizing

Für das Sessionizing muss eine Heuristik angewendet werden, da kein eindeutiger Identifikationsmechanismus zur Verfügung stand. Ein Request bzw. Logeintrag wird zu einer Session gezählt, wenn IP-Adresse sowie Agent mit denen der anderen Einträge identisch sind und der Zeitpunkt des Zugriffs sich innerhalb eines 30 Minuten-Intervalls zum ersten Request der Session befindet. Eine maximale Sessionlänge von 30 Minuten anzunehmen, hat sich in der Praxis als eine brauchbare Lösung erwiesen (siehe Spiliopoulou u. a. 2003).

Generalisierung der Zugriffe

Im Anschluss an das Sessionizing werden die Requests entsprechend der definierten Konzepthierarchien (siehe Abbildung 5.1 sowie Tabellen 5.1, 5.2 und 5.3) anhand ihres Logeintrags generalisiert. Von allen Zugriffen, die nicht auf diese Weise klassifiziert werden können, werden ca. 25% der entsprechenden Portalseiten manuell durch einem Experten begutachtet und in Seiten der Typen M, MN (stellvertretend

für MNI, MNE und MNINE) und N (stellvertretend für NI, NE und NINE) klassifiziert. Die Link-zu-Text-Verhältnisse in diesen Seiten werden mittels eines Softwareskripts ausgelesen („Text“ bezieht sich hierbei auf die Anzahl an Worten). Menüs sowie Kopf- und Fußzeilen werden dabei nicht berücksichtigt, da diese unabhängig vom Typ der Seite immer vorhanden sind. Aus diesen ermittelten Werten werden Klassifikationsregeln für die verschiedenen Seitentypen mittels des im Softwaretool WEKA (Witten u. Frank 2005) implementierten J48-Klassifikationsalgorithmus² generiert. Tabelle 6.2 enthält die Link-zu-Text-Verhältnisse der verschiedenen Seitentypen für die betrachteten Portale. Die Werte stimmen für alle Portale fast miteinander überein, d.h. die durch den Experten vorgenommene Klassifikation ist weitestgehend konsistent über alle Portale.

Tabelle 6.2: Hyperlink-zu-Text-Verhältnisse der verschiedenen Seitentypen

Portal	M	NI, NE, NINE	MNI, MNE, MNINE
P1	kleiner 1 zu 116 ($\leq 0,0086$)	größer 1 zu 10 ($> 0,0963$)	$> 0,0086$ und $\leq 0,0963$
P2	kleiner 1 zu 117 ($\leq 0,0085$)	größer 1 zu 10 ($> 0,0923$)	$> 0,0085$ und $\leq 0,0923$
P3	kleiner 1 zu 110 ($\leq 0,0091$)	größer 1 zu 11 ($> 0,0901$)	$> 0,0091$ und $\leq 0,0901$
P4	kleiner 1 zu 110 ($\leq 0,009$)	größer 1 zu 8 ($> 0,1261$)	$> 0,009$ und $\leq 0,1261$

Das Ergebnis der Klassifikationsregelgenerierung mit J48 für P2 ist in Abbildung 6.2 dargestellt. Die Ausgabe enthält Informationen zu den abgeleiteten Klassifikationsregeln (Abschnitt *Classifier model*), zu korrekt bzw. inkorrekt klassifizierten Seiten (Abschnitte *Stratified cross-validation* und *Summary*), zur Klassifikationsgüte auf Klassenebene (Abschnitt *Detailed Accuracy By Class*) sowie eine Konfusionsmatrix, die anzeigt, wie viele und welche Objekte falsch klassifiziert wurden (Abschnitt *Confusion Matrix*).

Das Ergebnis für P2 zeigt insgesamt eine gute Klassifikationsgüte: Der Klassifikationsfehler liegt bei ca. 4%; der Wert der *Kappa statistic*, die *True Positive* (TP)-, *Precision*-, *Recall*- und *F-Measure*-Werte liegen nahe bei eins und der *False Positive* (FP)-Wert nahe bei Null (siehe Tabelle 6.3 für eine Erläuterung dieser Kennzahlen); aus der Konfusionsmatrix ist erkennbar, dass der Klassifikationsfehler hauptsächlich darauf basiert, dass nicht immer richtig zwischen Seiten des Typs N und des Typs MN unterschieden werden konnte.

Die Fehlerkennzahlen *Mean absolute error*, *Root mean squared error* usw. sind für die Einschätzung der Klassifikationsgüte nicht von besonderer Relevanz und werden hier nicht weiter berücksichtigt.

²Der J48-Algorithmus ist eine adaptierte Version des C4.5 Klassifikationsalgorithmus.

```

==== Run information ===

Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:    page_type
Instances:   455
Attributes:  2
              12w
              class
Test mode:   10-fold cross-validation

==== Classifier model (full training set) ===

J48 pruned tree
-----
12w <= 0.0923
|   12w <= 0.0085: M (62.0)
|   12w > 0.0085: MN (223.0/2.0)
12w > 0.0923: N (170.0/11.0)

Number of Leaves : 3

Size of the tree : 5

Time taken to build model: 0.09 seconds

==== Stratified cross-validation ===
==== Summary ===

Correctly Classified Instances      436          95.8242 %
Incorrectly Classified Instances   19           4.1758 %
Kappa statistic                   0.9303
Mean absolute error               0.0436
Root mean squared error           0.1589
Relative absolute error           10.9532 %
Root relative squared error      35.6321 %
Total Number of Instances         455

==== Detailed Accuracy By Class ===

      TP Rate   FP Rate   Precision   Recall   F-Measure   Class
        0.984      0          1       0.984      0.992      M
        0.94       0.022      0.978      0.94       0.958      MN
        0.975      0.048      0.918      0.975      0.946      N

==== Confusion Matrix ===

  a   b   c   <- classified as
61   1   0 |   a = M
  0 218 14 |   b = MN
  0   4 157 |   c = N

```

Abbildung 6.2: Output des J48-Algorithmus für Portal P2

Tabelle 6.3: Klassifikationsgütekennzahlen

Kennzahl	Beschreibung
Kappa statistic	Übereinstimmung der prognostizierten Klassifikation von Objekten mit deren echter Klassenzugehörigkeit. Je näher der Wert bei 1 liegt, desto besser ist die Klassifikationsgüte.
True-Positive-Rate	Der Anteil an korrekt zu einer Klasse zugeordneten Objekten im Verhältnis zu allen zu dieser Klasse gehörenden Objekten in der Grundgesamtheit. Je näher die TP-Rate bei 1 liegt, desto höher ist der Anteil an korrekt klassifizierten Objekten.
False-Positive-Rate	Der Anteil an Objekten, die zu einer Klasse x falsch zugeordnet wurden, im Verhältnis zu allen Objekten, die nicht zur Klasse x gehören. Je näher die FP-Rate bei 0 liegt, desto geringer ist der Anteil an falsch klassifizierten Objekten.
Precision	Anteil der korrekt klassifizierten Objekte in der Menge aller Objekte, die einer Klasse zugeordnet werden. Je näher dieser Wert bei 1 liegt, desto größer ist der Anteil an korrekt klassifizierten Objekten in dieser Menge.
Recall	Bei J48 equivalent zur TP-Rate.
F-Measure	Harmonisches Mittel von Precision und Recall: $F = 2 * \text{Precision} * \text{Recall} / (\text{Precision} + \text{Recall})$. Der F-Wert ist ein Klassifikationsgüteindikator, der auf beiden Kennzahlen basiert. Je näher der Wert bei 1 liegt, desto besser ist die Gesamt-Klassifikationsgüte.

Bei den verbliebenen 75% der nicht-klassifizierten Seitenzugriffe wird durch Auslesen der entsprechenden HTML-Dateien das Hyperlink-zu-Text-Verhältnis ermittelt. Anhand der generierten Klassifikationsregeln wird die jeweilige Seite dann klassifiziert. Durch auswerten der Hyperlinks der N- und MN-Seiten kann im Anschluss ihr endgültiger Typ (NI, NE, NINE, MNI, MNE oder MNINE) bestimmt werden.

Analyse der Klickpfade bzw. Nutzungsmuster

Zur Analyse der Klickpfade werden die vorverarbeiteten und generalisierten Logeinträge mittels des Sequenzanalysetools *Web Utilization Miner* (WUM)³ ausgewertet. WUM (Spiliopoulou u. Faulstich 1998) erlaubt durch den Einsatz einer eigenen Abfragesprache (MINT) die Suche nach ganz bestimmten Sequenzen und Seitentypen. Daneben ermöglicht WUM für einen schnellen Überblick eine visuelle Darstellung der extrahierten Pfade. Beispielhaft sei die Abfrage für Klickpfade mit Zugriffen auf Seiten des Typs M in Abbildung 6.3 angegeben. Die Abfragen für andere Zielseitentypen bzw. für herunterladbare Dokumente sind analog aufgebaut. Aus dem Ergebnis einer WUM-Abfrage lassen sich die Effektivitäts- und Effizienzwerte (Definitionen 10 und 11 in Abschnitt 5.2.8) der Zugriffe ableiten.

³Eine für nicht-kommerzielle Zwecke frei nutzbare Version ist von www.hypknowsys.org herunterladbar.

```

select t
from node as a b, template a * b as t
where a.url = "H"
and wildcard.b.url != "L"
and wildcard.b.url !contains "DBL"
and wildcard.b.url !contains "@S"
and wildcard.b.url !contains "@GS"
and wildcard.b.url !contains "@DBS"
and wildcard.b.url != "M"
and b.url = "M"

```

Abbildung 6.3: MINT-Abfrage zur Extraktion der von der Homepage (Typ H) zu Seiten des Typs M verlaufenden Klickpfade

6.2.2 Ergebnisse und Interpretation

Für die Ergebnisherleitung wird die Präferenzstruktur des Entscheidungsträgers ermittelt und anschließend die Entscheidungsmatrix konstruiert.

Präferenzstruktur des Entscheidungsträgers

Die Gewichtung der Zielkriterien in der Entscheidungsmatrix muss den Präferenzvorstellungen des Entscheiders entsprechend vorgenommen werden. Zu diesem Zweck wird dieser gebeten, die Wichtigkeit von Zugriffen auf die verschiedenen hier betrachteten Seitentypen sowie auf herunterladbare Dokumente mittels einer Ratingskala von 1–5 anzugeben. Um daraus Gewichte ableiten zu können, muss diese hier als intervallbasierte Skala interpretiert werden. Eine entsprechende Annahme kann, wie oben beschrieben (Abschnitt 5.2.8), getroffen werden. Der Gewichtungsfaktor für einen Seitentyp bzw. den Typ DOC wird dann aus dem Verhältnis von Ratingwert dieses Typs zur Summe aller Ratingwerte ermittelt. Um sicher zu gehen, dass die so abgeleiteten Gewichte den Präferenzvorstellungen des Entscheidungsträgers entsprechen, werden ihm diese zur Überprüfung vorgelegt und ggf. korrigiert. Die ermittelten Ratings und Gewichte für P2 sind in Tabelle 6.4 abgebildet.

Die Ermittlung der Gewichte per Ratingskala erlaubt außerdem, den Stellenwert der betrachteten Seitentypen aus Sicht des Entscheidungsträgers zu erfassen. Das Ratingniveau bewegt sich hier zwischen „4“ und „5“, was darauf hindeutet, dass Zugriffe auf Seiten der hier betrachteten Typen sowie auf herunterladbare Dokumente als sehr wichtig angesehen werden. Die aus diesen Zugriffen abgeleiteten Effektivitäts- und Effizienzwerte können daher mit Berechtigung als Erfolgssindikatoren bzw. Kriterien in der Entscheidungsmatrix verwendet werden.

Neben der Erfassung der Präferenzen zu den Seitentypen und zum Typ DOC, muss erfragt werden, welche Präferenz der Entscheider bezüglich Effektivität und Effizienz von Zugriffen hat. Diese markiert er auf einer bipolaren Skala. Auch hier werden unter der Annahme, dass die Abstände zwischen den Skalenpunkten gleich sind, entsprechende Gewichte abgeleitet und dem Entscheider zur Überprüfung vorgelegt. Wie in Abbildung 6.4 ersichtlich, ist für den Betreiber von P2 in erster

Tabelle 6.4: Präferenzen zu den Seitentypen und zum Typ DOC

Typ	Rating (1–5)	Gewichtung
M	5	0.208
MNI	5	0.208
MNE	4	0.167
MNINE	5	0.208
DOC	5	0.208
Σ	24	~1

Linie wichtig, dass überhaupt Zugriffe auf die Zielseiten erfolgen. Ihm ist weniger wichtig, dass diese Seiten auf möglichst kurzen Pfaden erreicht werden.

Nach Herleitung der Gewichte, kann die Entscheidungsmatrix konstruiert werden.

Effektivität	○ ● ○ ○ ○ ○ ○ ○	Effizienz
\Rightarrow Gewichtung Effektivität: 0,8333		
\Rightarrow Gewichtung Effizienz: 0,1667		

Abbildung 6.4: Präferenz bezüglich Effektivität und Effizienz von Zugriffen

Entscheidungsmatrix

Zunächst werden einige zusätzliche Statistiken betrachtet, die bei der Interpretation der Ergebnisse hilfreich sein können. Aus Tabelle 6.5 lässt sich entnehmen, dass 76% aller Zugriffe auf normale Portalseiten (Typ P) erfolgten. Im Vergleich zu den anderen Portalen ist dieser Anteil relativ hoch, was auf den hohen Stellenwert normaler Seiten für die Besucher von P2 hinweist. Im Gegensatz dazu ist der Anteil an Zugriffen auf herunterladbare Dokumente (Typ DOC) mit 6% eher gering.

Tabelle 6.5: Häufigkeiten und Anteile von Service-Anfragen

Portal	Seite – Typ P	Herunterladbares Dokument – Typ DOC	
		Dokument	Typ DOC
P1	3632 (71,09%)	508 (9,94%)	
P2	11956 (76,02%)	934 (5,93%)	
P3	15689 (61,96%)	17 (0,06%)	
P4	5287 (32,26%)	488 (2,97%)	

Aus den Häufigkeiten der Seitenanfragen lässt sich schließen, dass reine Content-Seiten (Typ M) sowie gemischte Seiten (Typen MNI, MNINE) zu ungefähr gleich großen Teilen angefragt wurden (Tabelle 6.6). Im Vergleich zu den anderen Portalen sind die Anteile an Navigations-Seiten-Anfragen der Typen NI (ca. 10% aller Seitenanfragen) und NINE (ca. 30% aller Seitenanfragen) relativ groß (Tabelle 6.7), was auf eine tiefere Hierarchiestruktur auf P2 hinweist. Ein weiterer Hinweis

darauf sind die durchschnittlichen Pfadlängen, welche im Vergleich zu den anderen Portalen zu allen Zielseitentypen länger sind (Tabelle 6.8).

Tabelle 6.6: Häufigkeiten von Zugriffen auf Content-Seiten

Portal	H	M	MNI	MNE	MNINE
P1	964 (18,86%)	155 (3,03%)	501 (9,80%)	90 (1,76%)	1431 (28,00%)
P2	2086 (13,26%)	1051 (6,68%)	1157 (7,35%)	218 (1,38%)	985 (6,26%)
P3	4680 (18,48%)	42 (0,16%)	192 (0,75%)	44 (0,17%)	4559 (18,00%)
P4	2409 (14,69%)	519 (3,16%)	1196 (7,29%)	13 (0,07%)	569 (3,47%)

Tabelle 6.7: Häufigkeiten von Zugriffen auf Navigations-Seiten

Portal	NI	NE	NINE
P1	78 (1,52%)	15 (0,29%)	303 (5,93%)
P2	1589 (10,10%)	38 (0,24%)	4626 (29,41%)
P3	1855 (7,32%)	19 (0,07%)	4138 (16,34%)
P4	265 (1,61%)	1 (< 0,01%)	113 (0,68%)

Tabelle 6.8: Durchschnittliche Pfadlängen in den H-Sessions

Portal	M	MNI	MNE	MNINE	DOC
P1	4,9	2,9	4,4	1,3	4,2
P2	5,1	4,8	6,5	4,3	7,3
P3	4,1	4,1	5,9	2,4	4,3
P4	3,7	2,2	16,7	3,0	4,8

Die resultierende Entscheidungsmatrix zeigt, dass P2 auf Basis der gewählten Kriterien und Gewichte das zweitbeste Ergebnis erzielt (Tabelle 6.9). Das auf reiner Navigation (keine Nutzung von Suchmaschinen und Datenbanken) basierende Nutzerverhalten entspricht weitgehend den geäußerten Präferenzen: Die Anteile (Effektivitätswerte) der Seitentypen M, MNI und MNINE in den Sessions sind übereinstimmend mit der gleichen Gewichtung zu annähernd gleich großen Teilen vorhanden; der Anteil an MNE-Seiten ist geringer, aber auch weniger wichtig; deutlich geringer und der Gewichtung (und damit der Präferenz des Entscheiders) widersprechend ist der Anteil an Zugriffen auf herunterladbare Dokumente.

Tabelle 6.9: Entscheidungsmatrix zum Nutzerverhalten

w_j	0,208 M		0,208 MNI		0,167 MNE		0,208 MNINE		0,208 DOC		$U_{\text{Nutzerv.}}^{\text{SAW}}$
c_j	w_{jk}	c_{jk}	eff	efz	eff	efz	eff	efz	eff	efz	
P1	0,1126	0,2054	0,2815	0,3500	0,0871	0,2281	0,6434	0,7643	0,2399	0,2377	0,295
P2	0,1407	0,1975	0,1850	0,2067	0,0554	0,1544	0,1974	0,2321	0,0946	0,1376	0,147
P4	0,0993	0,2718	0,1902	0,4470	0,0054	0,0600	0,1306	0,3380	0,0289	0,2069	0,125
P3	0,0058	0,2455	0,0254	0,2459	0,0062	0,1686	0,3372	0,4190	0,0017	0,2353	0,110

Die Werte für die Gewichtungsfaktoren sind gerundet.

Tabelle 6.5 zeigt, dass ca. 6% aller Requests Anfragen nach herunterladbaren Dokumenten waren. Ein Vergleich mit Tabelle 6.6 offenbart, dass es – in absoluten Zahlen – jeweils annähernd gleich viele Anfragen (ca. 1000) nach Seiten der Typen M, MNI und MNINE sowie nach herunterladbaren Dokumenten (Typ DOC) gab. Dennoch sind in der Entscheidungsmatrix die Effektivitätswerte für Zugriffe auf diese Seiten bis zu mehr als doppelt so hoch, als die auf herunterladbare Dokumente. Das deutet darauf hin, dass DOC-Zugriffe in vergleichsweise weniger Sessions getätigt wurden, in diesen dann aber mehrmals Erfolgen konnten. Herunterladbare Dokumente waren nach diesem Ergebnis also eher nur für einen kleineren Teil der Nutzer interessant.

Zum besseren Verständnis der ermittelten Ergebnisse wird zusätzlich eine manuelle stichprobenartige Durchsicht von Seiten auf P2 vorgenommen – eine vollständige Durchsicht ist aufgrund des mehrere tausend Seiten umfassenden Contents im gesetzten Zeitrahmen der Untersuchung nicht realisierbar. Das Ergebnis der manuellen Betrachtung lässt darauf schließen, dass an vielen Stellen zwar über MNE- und MNINE-Seiten auf herunterladbare Dokumente verwiesen wurde, diese sich jedoch auf anderen Websites befanden. Herunterladbare Dokumente wurden somit oft nicht direkt von P2 geholt. Die Erhöhung der Anzahl an direkt von P2 herunterladbaren Dokumenten wäre ein Ansatz, der zu einem Nutzerverhalten führen könnte, dass eher den geäußerten Präferenzen entspräche. Das würde aber auch bedeuten, dass oft an Stelle einer Verlinkung zu einer Datei auf einer anderen Website, diese reproduziert und direkt auf P2 angeboten werden müsste. In den meisten Fällen wäre das aber nicht sinnvoll oder gar nicht möglich. Wenn die Anzahl der auf P2 direkt downloadbaren Dokumente nicht anders wesentlich erhöht werden kann oder soll, dann sind hier möglicherweise die Seitentypen M, MNI und MNINE – die entsprechende Verlinkungen zu Dateien auf anderen Websites enthalten können – in ihrer Wichtigkeit im Vergleich zu herunterladbaren Dokumenten unterbewertet. In diesem Fall wären die Präferenzen zu Überdenken.

Im Vergleich zu den anderen Portalen bewegen sich die Effektivitätswerte für Seiten des Typs M im oberen und für Seiten der Typen MNI und MNINE im mittleren Bereich. Insbesondere der Anteil an M-Seiten ist mit 14% ($eff(M)$ -Wert in der Matrix) der aktiven Sessions deutlich höher als bei den anderen Portalen. Dies zeigt, dass reine Content-Seiten auf P2 von den Nutzern häufig nachgefragt wurden und weist somit auf ein relativ hohes Interesse der Nutzer an diesen Seiten hin.

Allgemein fällt auf, dass bis auf Zugriffe auf Seiten des Typs MNE alle Effizienzwerte (Werte in den efz -Spalten) sich im unteren Bereich bewegen, d. h. die Pfade zu den entsprechenden Zielseiten waren in den H-Sessions in der Regel länger als auf den anderen Portalen. Besonders auffällig war das bei Content-Navigations-Seiten der Typen MNI und MNINE. In Tabelle 6.10 sind als Beispiel die kompletten Pfade von H- zu MNINE-Seiten abgebildet, die in mindestens zwei Sessions vorkamen und bei denen die MNINE-Seite zum ersten Mal in der jeweiligen Session aufgerufen wurde. Die erste Zahl in den Klammern hinter dem Seitentyp gibt an, wie oft dieser in der Session aufgerufen wurde (hier bei MNINE-Seiten also 1).

Die zweite Zahl gibt an, in wie vielen Sessions der Teilstiel von der H-Seite bis zu diesem Seitentyp vorkam. Es ist hier deutlich zu erkennen, dass die Pfade sehr viele reine Navigations-Seiten (Typen NI und NINE) enthielten.

Tabelle 6.10: Häufig vorkommende Pfadmuster für Zugriffe auf MNINE-Seiten

Muster	Pfad
1	H(1;2168) → NINE(1;822) → NINE(2;569) → NINE(3;235) → NINE(4;75) → NINE(5;23) → NINE(6 ;8) → MNINE(1;2)
2	H(1;2168) → NINE(1;822) → NINE(2;569) → NINE(3;235) → NINE(4;75) → NI(1;6) → MNINE(1;2)
3	H(1;2168) → NINE(1;822) → NINE(2;569) → NINE(3;235) → NINE(4;75) → MNI- NE(1;7)
4	H(1;2168) → NINE(1;822) → NINE(2;569) → NINE(3;235) → NI(1;45) → NI(2;19) → MNINE(1;2)
5	H(1;2168) → NINE(1;822) → NINE(2;569) → NINE(3;235) → NI(1;45) → MNINE(1;5)
6	H(1;2168) → NINE(1;822) → NINE(2;569) → NINE(3;235) → MNINE(1;24)
7	H(1;2168) → NINE(1;822) → NINE(2;569) → NINE(3;235) → M(1;3) → MNINE(1;2)
8	H(1;2168) → NINE(1;822) → NINE(2;569) → NI(1;88) → NI(2;20) → MNINE(1;5)
9	H(1;2168) → NINE(1;822) → NINE(2;569) → NI(1;88) → MNINE(1;7)
10	H(1;2168) → NINE(1;822) → NINE(2;569) → NI(1;88) → NINE(3;20) → NINE(4;10) → MNINE(1;4)
11	H(1;2168) → NINE(1;822) → NINE(2;569) → MNIN E(1;41)
12	H(1;2168) → NINE(1;822) → NINE(2;569) → MNI(1;36) → MNINE(1;6)
13	H(1;2168) → NINE(1;822) → NINE(2;569) → MNI(1;36) → NINE(3;11) → MNINE(1;2)
14	H(1;2168) → NINE(1;822) → NINE(2;569) → MNE(1;14) → MNINE(1;2)
15	H(1;2168) → NINE(1;822) → NI(1;80) → NI(2;18) → MNINE(1;4)
16	H(1;2168) → NINE(1;822) → NI(1;80) → MNINE(1;7)
17	H(1;2168) → NINE(1;822) → NI(1;80) → NINE(2;22) → NINE(3;15) → MNINE(1;4)
18	H(1;2168) → NINE(1;822) → COL(1;7) → MNINE(1;3)
19	H(1;2168) → NINE(1;822) → MNINE(1;31)
20	H(1;2168) → MNINE(1;74)
21	H(1;2168) → M(1;82) → MNINE(1;7)
22	H(1;2168) → M(1;82) → MNI(1;5) → MNINE(1;3)
23	H(1;2168) → M(1;82) → NINE(1;18) → NINE(2;9) → NINE(3;3) → NINE(4;3) → MNI- NE(1;2)
24	H(1;2168) → M(1;82) → NINE(1;18) → NINE(2;9) → MNINE(1;2)
25	H(1;2168) → MNI(1;60) → MNINE(1;2)
26	H(1;2168) → MNI(1;60) → NINE(1;9) → NINE(2;6) → MNINE(1;2)
27	H(1;2168) → H(2;106) → NINE(1;32) → MNINE(1;3)
28	H(1;2168) → H(2;106) → MNINE(1;2)
29	H(1;2168) → H(2;106) → MNI(1;22) → MNI(2;3) → MNINE(1;2)
30	H(1;2168) → H(2;106) → MNI(1;22) → MNINE(1;3)

Im Vergleich zu den anderen Portalen ist der prozentuale Anteil $n(\mathbf{X} = \mathbf{x})/N * 100$ an Sessions mit z. B. $\mathbf{x} = 1, 2, \dots, 5$ aufeinander folgenden NINE-Seitenabrufen deutlich größer (Abbildung 6.5). Gleicher gilt für NI-Seitenabrufe (hier nicht abgebildet). Eine manuelle, stichprobentartige Überprüfung lässt auf die Ursache dafür schließen: Viele NINE- und NI-Seiten enthalten nur sehr wenige Links (drei bis fünf) und sonst nichts; um zu den gewünschten Content-Seiten zu gelangen müssen daher entsprechend viele Navigations-Seiten angeklickt werden.

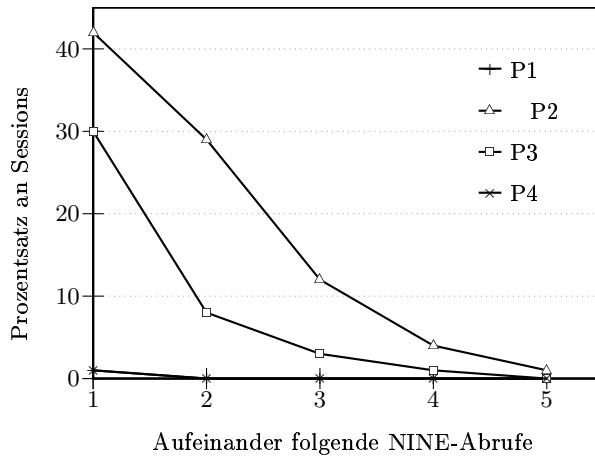


Abbildung 6.5: Anteile an Sessions mit aufeinander folgenden NINE-Abrufen

Da kurze Pfade den geäußerten Präferenzen (Abbildung 6.4) entsprechend nicht vorrangige Priorität haben, wirken sich die niedrigen Effizienzwerte nicht sehr negativ auf das Gesamtergebnis aus. Dennoch wäre hier zu überlegen, ob eine Verkürzung der Pfade nicht sinnvoll wäre, um den Nutzern ein komfortableres Surfen auf dem Portal zu ermöglichen. Ein Ansatz wäre, zumindest die direkt aufeinander folgenden Seiten mit sehr wenigen Links (weniger als fünf) zusammenzufassen. So würde die hierarchische Struktur des Portals abgeflacht und gleichzeitig gäbe es weniger Requests, wodurch das Serversystem entlastet würde. Möglich wäre es auch, reine Navigations-Seiten mit etwas Content zu befüllen, d.h. diese in Content-Navigations-Seiten (Typen MNI, MNE oder MNINE) umzuwandeln. Eine entsprechende Struktur liegt beispielsweise bei den Portalen P1 und P3 vor. Nutzer könnten dann durch die einem Link beigefügten Informationen besser einschätzen, ob es sich lohnt diesen anzuklicken oder nicht.

Zusammenfassend lässt sich sagen, dass P2 auf Basis der verwendeten Kriterien, der vorliegenden Präferenzstruktur und im Vergleich zu den anderen Portalen ein gutes Ergebnis aufweist. Schwerwiegende Schwachstellen wurden nicht aufgedeckt. Zu überdenken wäre die Präferenzvorstellung bezüglich herunterladbarer Dokumente. Ein möglicher Verbesserungsansatz wäre die Abflachung der Portalstruktur durch eine Überarbeitung an Stellen, wo mehrere reine Navigations-Seiten mit nur sehr wenigen Links direkt aufeinanderfolgen. Anzustreben wäre eine Struktur mit dem Aufbau *Homepage* → $[0, l]$ *Navigations-Seiten* (Typen NI, NINE) → $[0, m]$ *Content-Navigations-Seiten* (Typen MNI, MNINE) → $[0, n]$ *Content-Seiten* (Typ M) oder *Downloaddokumente* (Typ DOC) mit $l, m, n \in \mathbb{N}$. Ideal wäre hierbei ein Gleichgewicht zwischen nicht zu hoher Anzahl an Hierarchieebenen und nicht zu sehr gefüllten bzw. unübersichtlichen Seiten.

6.2.3 Diskussion

Ziel der Analyse des Nutzerverhaltens im hier vorgestellten Ansatz ist eine Überprüfung, inwieweit dieses in Bezug auf die angefragten Seitentypen und herunterladbaren Dokumente sowie die dafür benötigten durchschnittlichen Klickpfadlängen den Präferenzvorstellungen des Entscheiders bzw. Portalanbieters entspricht. Als Bewertungskriterien werden die Effektivität (Definition 10) sowie die durchschnittliche Effizienz (Definition 11) der Zugriffe herangezogen. Laut Proposition 1 lassen sich anhand dieser Kriterien Aussagen zum entgegengebrachten Interesse für einen Seitentyp (bzw. für herunterladbare Dokumente) sowie zu Schwachstellen in der Menü- und Portalstruktur machen.

Die Ergebnisse der empirischen Auswertung des Nutzerverhaltens deuten daraufhin, dass Proposition 1 Gültigkeit besitzt. Die Identifikation von Schwachstellen erfolgt allerdings nur aus der subjektiven Perspektive des Entscheiders bzw. Portalanbieters bezüglich des gewünschten Nutzerverhaltens. Einen objektiven Maßstab gibt es dafür nicht.

Durch den Vergleich mit anderen Portalen ist recht gut analysierbar, welche Strukturierung von Seiten ein bestimmtes Nutzerverhalten hervorruft. Die Identifizierung von *good* bzw. *best practices* sollte hier daher gut möglich sein.

In der vorliegenden Untersuchung ist der Fokus auf das gesamte Portal gerichtet. Die Ergebnisse beziehen sich dementsprechend auf alle Seiten die im Untersuchungszeitraum angeklickt wurden. Wenn ein Portal in verschiedenen Bereichen (z.B. Rubriken) stark voneinander abweichende Strukturen aufweist, bilden die Ergebnisse für Effektivität und Effizienz eine Art „Mittelwert“ über alle Portalbereiche. Diese Werte hätten dann eine geringere Aussagekraft. Bei solchen Portalen wäre es daher sinnvoll, homogene Teilbereiche separat zu untersuchen. Die Portale in der vorliegenden Studie sind von ihrem Gesamtaufbau überwiegend homogen. Man kann somit davon ausgehen, dass die Ergebnisse relativ hohe Aussagekraft haben.

Die Genauigkeit der Ergebnisse hängt des Weiteren von der genutzten Datenbasis ab. Ein grundsätzliches Problem der Analyse von Nutzerverhalten auf Basis von Webserver-Logdaten ist die nicht-vollständige Rekonstruktion von Sessions (siehe Abschnitt 3.3.2). Nicht alle Aktionen, die ein Nutzer am Browser ausführt (z.B. drücken des „Zurück“-Knopfes), werden im Log eines Webservers festgehalten. Daneben können ohne eindeutige Identifizierung einzelne Requests bei der Rekonstruktion von Sessions fehlerhaft zugeordnet werden. Die Qualität der Ergebnisse aller Folgeanalysen hängt von der Korrektheit der Sessionrekonstruktion ab. Das hier eingesetzte Verfahren mit Vergleich von IP-Adresse und Agent sowie einem 30 Minuten Sessionlimit ist laut bestehender Untersuchungen (Spiliopoulou u. a. 2003) relativ zuverlässig und das beste bei Fehlen eindeutiger Identifizierungsmechanismen, wie z.B. Cookies.

Für eine regelmäßige Anwendung in der Praxis ist die Abstrahierung der Requests durch Klassifikation der entsprechenden Seiten anhand des Verhältnisses von Hyperlinks zu Text allerdings eher unpraktikabel. Dieser Prozess ist aufwän-

dig und außerdem können beim automatisierten Auslesen der Portalseiten Fehler auftreten, da nicht alle Seiten die gleiche interne Struktur aufweisen. Wenn z. B. im Content Management System Änderungen an der Vorlage zum Aufbau von Seiten vorgenommen werden, dann sind anschließend neu erstellte Seiten anders strukturiert, als die bereits bestehenden. Kopf- und Fußzeilen können beispielsweise anders definiert sein. Eine manuelle Kontrolle der klassifizierten Seiten war bei den betrachteten Portalen daher immer notwendig. Außerdem werden multimediale Inhalte bei der Betrachtung von Link-zu-Text-Verhältnis nicht berücksichtigt.

Für eine zukünftige, regelmäßige Evaluation wäre es vorteilhafter, bei der Erstellung neuer Portalseiten eine entsprechende Information zum Typ einer Seite mitzugeben. Das könnte z. B. mittels eines *HTML-Meta-Tags* „Seitentyp“ realisiert werden. Der Inhalt dieses Tags könnte dann bei der Auswertung der Seite einfach ausgelesen werden. Damit würde eine Klassifikation anhand des Link-zu-Text-Verhältnisses überflüssig und Fehler würden vermieden.

Zu beachten ist, dass die Pfadanalyse nur zeigt, ob sich die Nutzer den Präferenzen des Portalanbieters entsprechend verhalten. Die Ergebnisse sagen nichts über deren Zufriedenheit mit dem Portal und den Inhalten aus. Dazu müssen die Nutzermeinungen erfragt werden.

6.3 Bewertung der Nutzermeinungen

Zur Erfassung der Nutzermeinungen wurden anonymen Online-Umfragen auf den Portalen durchgeführt. Die Ergebnisse werden zur Validierung des postulierten Pfadmodells (Abbildung 5.4) herangezogen. Anschließend folgt die Konstruktion der Entscheidungsmatrix auf Basis des validierten und – falls nötig – überarbeiteten Pfadmodells.

Für die Umfrage wurde auf den Homepages der Portale jeweils ein Link zum Online-Fragebogen gesetzt. Zusätzlich bekamen alle registrierten Portalnutzer per eMail eine Einladung zur Teilnahme. Diese beinhaltete eine kurze Erläuterung zum Zweck⁴ der Umfrage sowie den Link zum Fragebogen. Die Startseite zur Online-Umfrage enthielt eine ausführliche Einführung und es wurde erläutert, wie die Fragen mittels der Ratingskalen zu beantworten sind. Zusätzlich folgte der Hinweis, dass die Umfrage aus Datenschutzgründen anonym erhoben wird und damit die Privatsphäre der Teilnehmer gewahrt bleibt. Daneben wurde für eventuelle Rückfragen eine Kontaktadresse angegeben. Beabsichtigt war es, einen möglichst seriösen und vertrauenswürdigen Eindruck bei den Teilnehmern zu erzeugen.

In Abhängigkeit vom jeweiligen Portal betrug der Umfragezeitraum vier bis acht Wochen. Die Umfragen wurden im Dezember 2006 (P1, P2, P3) sowie im Januar 2007 (P4) durchgeführt. Es erfolgten insgesamt 2216 Zugriffe auf die Einführungsseite. Gestartet hatten die Umfrage 1049 Teilnehmer, aber nur 526 begannen

⁴Die Teilnahme der Portalnutzer wurde erbeten, um mit deren Hilfe eine Einschätzung der Portalqualität sowie Hinweise zu Schwachstellen zu erhalten.

dann tatsächlich mit dem Ausfüllen des Fragebogens. Vollständig ausgefüllt wurde der Fragebogen von 223 Befragten (Tabelle 6.11).

Tabelle 6.11: Statistik zur Umfrageteilnahme				
Portal	Kontakte	gestartet	ausgefüllt	vollständig
P1	779	297	145	63
P2	440	240	148	61
P3	566	245	88	32
P4	431	267	145	67
Σ	2216	1049	526	223

Unvollständigkeit von Erhebungsdaten aufgrund nicht komplett ausgefüllter Fragebögen ist in den Sozialwissenschaften ein bekanntes und verbreitetes Problem. King u. a. (2001) beispielsweise berichten von durchschnittlich 50% unvollständiger Fragebögen bei politischen Umfragen.

Statistische Auswertungstools arbeiten meist nur mit vollständigen Daten. In der Regel werden unvollständige Datensätze daher aus den zu untersuchenden Daten entfernt (engl.: *listwise deletion*), d. h. es werden nur die vollständigen Fälle betrachtet (engl.: *complete case analysis*). Im besten Fall verliert man durch diese Vorgehensweise wertvolle Informationen, da schon Datensätze mit nur einer oder wenigen nicht beantworteten Fragen unberücksichtigt bleiben. Im schlimmsten Fall kann dies zu systematisch mit Fehlern behafteten Ergebnissen (engl.: *non-response bias*) führen.

Alternative Ansätze simulieren fehlende Werte auf Basis der vorhandenen Daten. In der englischen Literatur wird hierfür der Begriff *imputation* verwendet (siehe Rässler u. Riphahn 2006). Untersuchungen haben gezeigt, dass eine Imputation fehlender Werte in den meisten Fällen zu besseren Ergebnissen führt, als die Betrachtung nur der vollständigen Fälle (z. B Heitjan u. Rubin 1990; King u. a. 2001).

Wie aus Tabelle 6.11 ersichtlich, sind ca. 58% (303 von 526) der ausgefüllten Fragebögen unvollständig. Um auch diese zu erschließen und einen möglichen Einfluss eines Non-Response-Bias zu minimieren, werden die Daten in der Aufbereitungsphase durch entsprechende Ergänzungsverfahren vervollständigt. Anschließend erfolgt die Validierung des Pfadmodells (Abbildung 5.4) und die Konstruktion der Entscheidungsmatrix.

6.3.1 Datenaufbereitung

Die Daten werden zunächst durch Imputation vervollständigt. Im Anschluss wird das Pfadmodell (Abbildung 5.4, Hypothesen in Tabelle 5.4) mittels der empirischen Ergebnisse validiert und ggf. respezifiziert. Anhand des validierten Modells lässt sich dann auf die Eignung der hier vorgeschlagenen Kriterien zur Erfassung der individuellen Auswirkungen (Abschnitt 5.2.8) schließen.

Imputation fehlender Werte

Zweck der Imputation ist es, vollständige Daten zu erhalten. Diese können dann mit jedem beliebigen statistischen Tool analysiert werden. Angestrebt wird eine möglichst exakte Bestimmung fehlender Werte, damit die auf Basis der vervollständigten Daten berechneten statistischen Schätzwerte die gleichen sind, als würden sie anhand der vollständigen Daten ermittelt. Die Multiple Imputation ist dafür ein geeigneter Ansatz (siehe z. B Heitjan u. Rubin 1990; King u. a. 2001) und wird hier angewendet – eine Einführung zur Thematik ist in Anhang A gegeben.

Zunächst ist die notwendige Mindestanzahl m an Imputationen zu bestimmen, um eine übermäßige Erhöhung der Standardabweichungen relativ zu $m \rightarrow \infty$ bei den Variablen mit fehlenden Werten zu vermeiden. Die größten Anteile an fehlenden Werten liegen bei allen Portalen für die Items VERT2, RISK1, RISK2 vor (siehe Tabelle 6.12). Im Verhältnis zur Gesamtmenge der Daten aus allen vier Portalen, bewegen sich diese Anteile zwischen 29% und 35%. Bei $m = 5$ Imputationen und einem 35%-gem Anteil fehlender Werte für eine Variable, erhöht sich deren Standardabweichung im Vergleich zu $m \rightarrow \infty$ um den Faktor 1,036 (Rubin 1987, S. 114, Tabelle 4.1). Da dieses ein tolerierbarer Wert ist wird $m = 5$ gesetzt.

Tabelle 6.12: Anteile fehlender Werte

Portal	Items							
	INFQ1	INFQ2	INFQ3	INFQ4	INFQ5	SYSQ1	SYSQ2	SYSQ3
P1	2%	2%	13,1%	5,5%	22,7%	1,3%	1,3%	0,6%
P2	2%	2%	12,8%	4,7%	25%	2%	1,3%	2,7%
P3	9%	3,4%	12,5%	4,5%	13,6%	1,1%	3,4%	1,1%
P4	3,4%	2%	12,4%	8,2%	13,1%	1,3%	1,3%	1,3%
\sum	3,6%	2,2%	12,7%	5,8%	19,2%	1,5%	1,7%	1,5%
	SYSQ4	SYSQ5	SYSQ6	ZUFR1	ZUFR2	WISS1	WISS2	VERT1
P1	7,5%	2,7%	13,1%	0%	3,4%	5,5%	9,6	8,9%
P2	5,4%	3,3%	16,2%	1,3%	4,7%	6%	11,4%	7,4%
P3	6,8%	6,8%	12,5%	0%	2,2%	6,8%	12,5%	10,2%
P4	4,1%	2,7%	8,9%	1,3%	6,2%	6,8%	12,4%	9,6%
\sum	5,8%	3,6%	12,7%	0,7%	4,3%	6,2%	11,4%	8,9%
	VERT2	RISK1	RISK2	LOYA1	LOYA2			
P1	28,2%	27,5%	33,1%	6,8%	12,4%			
P2	35,8%	29%	33,1%	5,4%	8,7%			
P3	46,5%	36,3%	43,1%	6,8%	11,3%			
P4	26,2%	25,5%	33,1%	7,5%	8,2%			
\sum	32,8%	28,8%	34,7%	6,6%	10%			

Für die Imputation werden die Erhebungsdaten zu den 21 Fragen (Items) exklusive der demographischen Fragen aus allen vier Portalumfragen zusammengeführt. Die Items des Fragebogens stellen hierbei die Variablen im Imputationsmodell dar. Eine zusätzlich eingeführte Variable PORT (Portal) dient als Identifizierer für die

Portalzugehörigkeit eines Datensatzes. Dadurch können bei der Imputation Unterschiede in den Verteilungen der Erhebungsdaten zwischen verschiedenen Portalen – sofern vorhanden – berücksichtigt werden. Für die anschließende Auswertung der Daten wird diese Variable nicht mehr benötigt.

Die angewendete Imputationsmethode basiert auf logistischer Regression. Hierbei wird ein logistisches Regressionsmodell für die jeweils betrachtete (abhängige) Variable in Abhängigkeit aller anderen (unabhängigen) Variablen, der so genannten Kovariaten, erzeugt. Basierend auf diesem Modell wird ein neues Regressionsmodell mittels der prognostizierten A-posteriori-Verteilung der Parameter simuliert und für die Imputation der fehlenden Werte bei der betrachteten Variable herangezogen (Rubin 1987, S. 169 f.).

Die logistische Regression wird üblicherweise bei dichotomen Variablen angewendet, ist aber auch zur Analyse kategorischer Variablen mit mehr als zwei Ausprägungen gut geeignet. Letzteres liegt hier vor. Die Variablen besitzen, basierend auf der hier verwendeten fünfstufigen Ratingskala, eine ordinale Skalierung mit fünf Ausprägungen.

Die Imputation auf Basis logistischer Regression benötigt ein monotonen Muster der fehlenden Werte in der betrachteten Datentabelle. Eine entsprechende Struktur ist gegeben, wenn bei einem Datensatz mit Y_1, Y_2, \dots, Y_k Variablen (in dieser Reihenfolge) für Variable Y_i der Wert fehlt und für alle Variablen Y_j dieses Datensatzes mit $j > i$ ebenfalls die Werte fehlen (Rubin 1987, S. 170 f.), siehe Beispiel in Tabelle 6.13.

Tabelle 6.13: Beispiel eines monotonen Musters fehlender Werte

Datensatz	Y_1	Y_2	Y_3
1	1	1	0
2	1	0	0
3	0	0	0

Ein zum großen Teil monotonen Muster fehlender Werte wird in der hier vorliegenden Datentabelle durch Umstellen der Reihenfolge der Variablen erzielt. Im Unterschied zur ursprünglichen Reihenfolge (siehe Tabelle 6.12) stehen die Variablen mit den größten Anteilen an Fehlwerten nun am Ende. Die neue Reihenfolge ist folgende: PORT, INFQ1, INFQ2, INFQ3, INFQ4, INFQ5, SYSQ1, SYSQ2, SYSQ3, SYSQ4, SYSQ5, SYSQ6, ZUFR1, ZUFR2, WISS1, WISS2, LOYA1, LOYA2, VERT1, VERT2, RISK1, RISK2.

Die fehlenden Werte, die nicht-monotone Fehlmuster bilden, müssen vor der eigentlichen Imputation ergänzt werden. Das hierfür geeignete Verfahren ist die Imputation mittels *Markov Chain Monte Carlo* (MCMC)-Simulation. MCMC wird angewendet, um Pseudo-Zufallsziehungen aus einer multidimensionalen, schwer erfassbaren Wahrscheinlichkeitsverteilung f mittels Markov-Ketten zu tätigen. Eine Markov-Kette besteht aus einer Sequenz von Zufallsvariablen, bei der die Verteilung einer Variablen nur auf dem Wert der Vorgängervariablen basiert. Die MCMC-Simulation konstruiert eine Markov-Kette, bei der sich die Verteilung der Elemente

mit jedem neuen Glied auf eine stationäre Verteilung zu bewegt. Diese würde im Fall einer unendlichen Anzahl von Gliedern der Verteilung von f entsprechen. Bei einer ausreichenden Länge der Kette ist jede zusätzlich gezogene Variable daher annähernd eine Zufallsziehung aus f (Schafer 1997, S. 68 f.).

Sowohl die Regressions- als auch die MCMC-Methode setzen eine multivariate Normalverteilung der zugrunde liegenden Daten voraus. In der Realität weicht die Verteilung in einer Untersuchungsstichprobe jedoch meist von der multivariaten Normalverteilung ab. Dennoch können bei einem nicht zu hohen Anteil fehlender Werte (<50%) robuste Ergebnisse erzielt werden (Schafer 1997, S. 211 f.). Das erklärt sich daraus, dass das Imputationsmodell nur auf die fehlenden Werte einer Variable angewendet wird und nicht auf die ganze Datentabelle (Schafer 1997, S. 147 f.). Zur Verdeutlichung dieses Sachverhalts sei folgendes, abstraktes Beispiel gegeben:

Beispiel 1 Der Anteil fehlender Werte einer Variable ist 30%. Die angewendete multiple Imputationsmethode produziert 80% valide und 20% defiziente Imputationswerte. Damit liegt der Anteil der Verfälschung durch die imputierten Daten bei 20% von 30%, also bei 6% aller Werte dieser Variable. Aus diesen Daten abgeleitete Analyseergebnisse sind damit zu 94% korrekt.

Die visuelle Darstellung der Erhebungsdaten in Histogrammen gibt einen ersten Hinweis, ob eine von der Normalverteilung abweichende Verteilung vorliegt. Beispielsweise seien hier die Histogramme für die Items bzw. Variablen INFQ3 und VERT2 in Abbildung 6.6 gezeigt (die Histogramme zu allen Variablen sind in Anhang B dargestellt). Diese repräsentieren die zwei in den hier vorliegenden Daten vorkommenden Hauptkategorien von Häufigkeitsverteilungen. Die Mehrheit der Variablen hat eine zu INFQ3 ähnliche Histogrammform, die auf eine nicht zu starke Abweichung von der Normalverteilung schließen lässt. Die Variablen VERT1, RISK1, RISK2, LOYA1 und LOYA2 haben zu VERT2 ähnliche Formen, welche auf eine deutliche Abweichung hinweisen. Die Anteile fehlender Werte bei diesen sechs Variablen liegen in der gesamten Datenmenge aus allen Portalen jeweils unter 35% (siehe Tabelle 6.12). Bei diesen recht geringen Anteilen kann davon ausgegangen werden, dass eine Imputation auf Basis der Normalverteilungsannahme der Daten zu relativ robusten Ergebnissen führt.

Im Anschluss an die Imputation erfolgt die Validierung des postulierten Pfadmodells (Abbildung 5.4) für jedes Portal separat.

Modellvalidierung

Zur Modellvalidierung wird der *Partial Least Squares* (PLS)-Ansatz (Chin 1998; Tennehaus u. a. 2005) verwendet. Dieser eignet sich insbesondere für Studien mit explorativem Charakter.

Die postulierten Abhängigkeiten im D/M92, welches Grundlage des hier entwickelten Modells ist, werden in verschiedenen Untersuchungen bestätigt (siehe

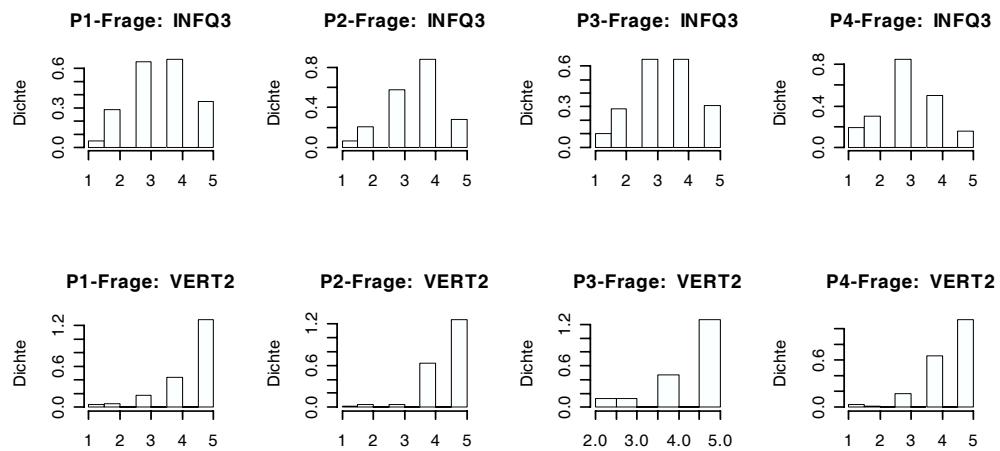


Abbildung 6.6: Histogramme zu den erhobenen Daten für die Items bzw. Variablen INFQ3 und VERT2 der Portale P1–P4

DeLone u. McLean 2003). Dem Autor ist allerdings keine entsprechende Studie bekannt, in der die hier vorgeschlagenen Kriterien zur Erfassung individueller Auswirkungen Anwendung finden. Aus diesem Blickpunkt handelt es sich zum gewissen Grad um eine explorative Studie, womit PLS als Validierungsmethode geeignet ist.

PLS ist relativ robust bei Daten, die von der multivariaten Normalverteilung abweichen (Gefen u. a. 2000). Das Verfahren kann hier daher eingesetzt werden, obwohl einige Variablen eine deutliche Abweichung in der Verteilung vorweisen (siehe Anhang B).

Die Stichprobengröße sollte für PLS mindestens 10-mal so hoch sein, wie die Anzahl der Messvariablen für das komplexeste Konstrukt (bzw. latente Variable), siehe Gefen u. a. (2000). Beim vorliegenden Pfadmodell ist das das Konstrukt SYSQ mit sechs Messvariablen (SYSQ1–SYSQ6), d. h. die notwendige Mindestanzahl an Untersuchungsfällen beträgt 60. Inklusive der imputierten Datensätze wird diese Mindestanzahl an Fällen bei jedem Portal überschritten (siehe Spalte „abgeschlossen“ in Tabelle 6.11). Die Validierung des Pfadmodells erfolgt für jede Portalstichprobe separat.

Zunächst erfolgt eine Prüfung auf Validität und Reliabilität. Bei der Validitätsprüfung werden zwei Eigenschaften des Modells betrachtet: die Konstruktvalidität und die strukturelle Validität.

Die Überprüfung der Konstruktvalidität (bzw. des Messmodells) zeigt, ob die zur Erfassung eines Konstrukt bzw. einer latenten Variable zugeordneten Indikator- bzw. Messvariablen (repräsentiert durch die Items bzw. Fragen des Fragebogens) sich eindeutig nur auf dieses beziehen (man spricht hier von „laden“) oder ob auch andere Konstrukte mit erfasst werden (hierfür wird der englische Begriff *cross loading* verwendet). Die Überprüfung erfolgt mittels einer Konvergenz- und einer Diskriminanzanalyse. Dazu werden bei PLS die t-Werte zu den Faktorladungen, die Faktorladungsmatrix sowie die Average Variance Extracted (AVE)-Werte heran-

gezogen (Gefen u. Straub 2005). Zu Details dieser Kennzahlen siehe Tabelle 6.14.

Tabelle 6.14: Indikatoren für Konstruktvalidität

Validitätsmaß	Kriterium	Heuristik
Konvergenzvalidität	t-Werte zu Faktorladungen	Faktorladungen sind signifikant bei einem $p \leq 0,05$.
Diskriminanzvalidität	Faktorladungen (Crossloadings)	Die Faktorladungen der zu einem Konstrukt zugeordneten Messvariablen sollten deutlich über den Faktorladungen der anderen diesem Konstrukt nicht zugeordneten Messvariablen liegen.
	AVE	Der AVE gibt den Anteil der durchschnittlichen Varianz in den Messvariablen an, welcher von deren zugehörigem Konstrukt erfasst wird, im Verhältnis zur Varianz der Messfehler. Der AVE ist definiert als $AVE = \frac{\sum \lambda_i^2}{(\sum \lambda_i^2) + (\sum (1 - \lambda_i^2))}$ mit λ_i als den Anteil der erfassten Varianz von Messvariable i . \sqrt{AVE} eines Konstruktions sollte „viel“ größer sein, als dessen Korrelation mit jedem anderen Konstrukt des Modells.

Die Überprüfung der strukturellen Validität (bzw. des Strukturmodells) zeigt, inwieweit die postulierten Abhängigkeiten zwischen den Konstrukten vorhanden sind. Dazu werden in PLS die t-Werte zu den Pfadkoeffizienten sowie das Bestimmtheitsmaß R^2 der endogenen Konstrukte herangezogen (Gefen u. a. 2000). Zu Details dieser Kennzahlen siehe Tabelle 6.15.

Tabelle 6.15: Indikatoren für strukturelle Validität

Kriterium	Heuristik
t-Werte zu Pfadkoeffizienten	Abhängigkeiten zwischen zwei Konstrukten sind signifikant für ein $p \leq 0,05$.
R^2	Je höher der Wert des Bestimmtheitsmaßes R^2 eines endogenen Konstruktions ist, desto größer ist der Anteil dessen Varianz, der durch die exogenen Konstrukte bestimmt wird.

Die Betrachtung der Reliabilität der Konstrukte zeigt deren Genauigkeit und Zuverlässigkeit bei der Datenerfassung. Es wird analysiert, inwieweit ein Umfrageteilnehmer unter gleich bleibenden Voraussetzungen dieselbe Frage wiederholt gleich oder annähernd gleich beantworten würde. Cronbachs α ist hierfür ein gebräuchlicher Indikator (Cronbach 1951). Dieser misst die durchschnittliche Inter-Korrelation unter den Messvariablen (Items) einer latenten Variable (Konstrukt) und ist definiert als

$$\alpha = \frac{N}{N - 1} \frac{\sigma_X^2 - \sum_{i=1}^N \sigma_{Y_i}^2}{\sigma_X^2}, \quad (6.1)$$

mit N als Anzahl der Messvariablen zur betrachteten latenten Variable, σ_X^2 der observierten Gesamtvarianz der zugehörigen Messvariablen und $\sigma_{Y_i}^2$ der Varianz von Messvariable i . Für gute Reliabilität sollte das α eines Konstrukt nicht unter 0,7 liegen. Bei explorativen Studien ist ein Wert von 0,6 akzeptabel (Gefen u. a. 2000, S.63).

Die verwendete Software ist SmartPLS 2.0 M3 (Ringle u. a. 2007). Zur Validierung mit dem PLS-Algorithmus werden die in SmartPLS voreingestellten Parameter⁵ übernommen. Für die Signifikanztests mittels der t-Statistik wird der in SmartPLS implementierte Algorithmus für Bootstrapping (Davison u. Hinkley 1999) verwendet. Vorgegebene Parameter sind (1) die Anzahl der Fälle (*cases*) und (2) die Anzahl der Simulationsdurchläufe (*samples*). Die Anzahl der Fälle entspricht dabei der jeweiligen Stichprobengröße eines Portals. Um verlässliche t-Werte zu erzielen, wird für die Anzahl der Simulationsdurchläufe das Zehnfache der Anzahl der Fälle gesetzt.

Die Validierung ergibt zunächst einen ungenügenden Fit des Pfadmodells. Die Messvariablen (Items bzw. Fragen) ZUFR1 und ZUFR2, die dem Konstrukt ZUFR (Zufriedenheit) zugeordnet sind, laden auch auf die Konstrukte SYSQ (Systemqualität) und INFQ (Informationsqualität). Es liegt also ein *cross loading* der ZUFR-Items auf die SYSQ- und INFQ-Konstrukte vor. Das zeigt, dass auf Basis der hier gewählten Messvariablen keine eindeutige Trennung zwischen Nutzerzufriedenheit und empfundener System- bzw. Informationsqualität vorgenommen werden kann. Die Messvariable INFQ5 hat nur eine sehr geringe Faktorladung auf ihr Konstrukt (INFQ) und ist damit nicht aussagekräftig. Die Faktorladungen der Messvariablen von SYSQ deuten darauf hin, dass dieses Konstrukt aus zwei Subkonstrukten besteht.

Im Rahmen der Respezifizierung des Modells werden das Konstrukt ZUFR sowie Item INFQ5 entfernt. Das Konstrukt SYSQ wird in die zwei Subkonstrukte PREQ (Präsentationsqualität; Items SYSQ1–SYSQ3) und TECQ (Technische Qualität; Items SYSQ4–SYSQ6) aufgespalten.

Die erneute Validierung nach diesen Änderungen ergibt einen guten Modellfit (siehe Anhang C). Es liegen jedoch nicht alle angenommenen Abhängigkeiten (Abbildung 5.4) zwischen den Konstrukten vor, und ermittelte signifikante Abhängigkeiten sind auch nicht bei allen Portalen identisch. Das respezifizierte Pfadmodell ist in Abbildung 6.7 dargestellt. Die Dicke der Pfeile zeigt an, bei wie vielen Portalen signifikante Abhängigkeiten vorliegen.

Das Validierungsergebnis zum respezifizierten Modell bestätigt die Hypothesen (Tabelle 5.4) H3, H4, H9, H10, H12, H13 sowie H14. Demnach wirken sich Präsentationsqualität (PREQ) und Informationsqualität (INFQ) sowohl auf die Wissens-

⁵(1) Weighting Scheme: Path Weighting Scheme; (2) Data Metric: Mean 0, Var 1; (3) Abort Criterion: 1.0E-5; (4) Initial Weights: 1.0.

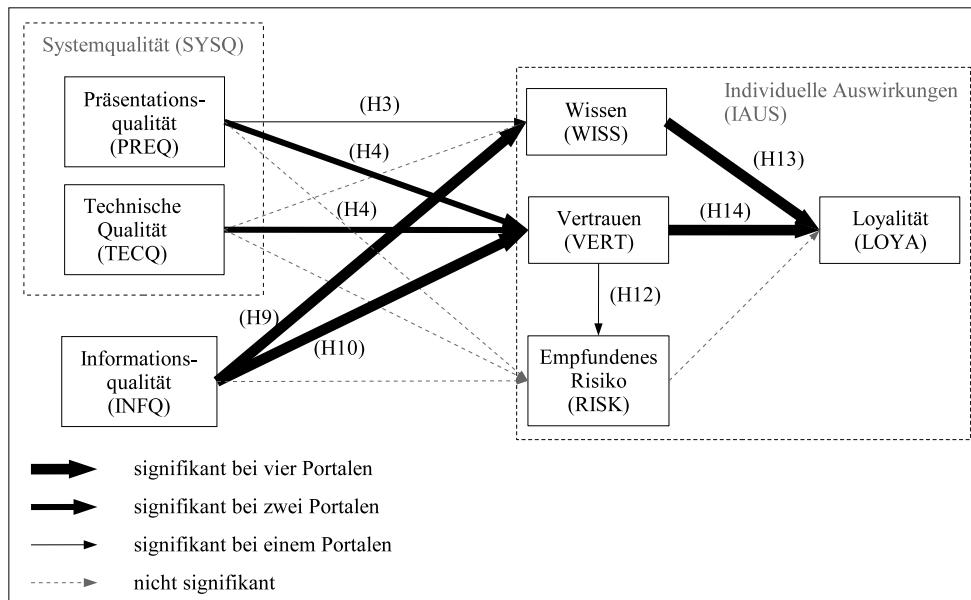


Abbildung 6.7: Respezifiziertes Pfadmodell zu den Nutzermeinungen

(WISS) als auch auf die Vertrauensbildung (VERT) aus. Die technische Qualität (TECQ) wirkt nur auf die Vertrauensbildung. Vertrauensbildung und Wissen wiederum beeinflussen die Loyalität (LOYA). Vertrauensbildung beeinflusst das empfundene Risiko (RISK). Die restlichen Hypothesen können hier aufgrund der empirischen Ergebnisse nicht bestätigt werden. Das Resultat entspricht den in Abschnitt 3.2.8 vorgestellten Studien zu Auswirkungen von System- und Informationsqualität.

Insgesamt deutet das Validierungsergebnis darauf, dass der in Proposition 2 vorgeschlagene, positive Zusammenhang zwischen Systemqualität und Informationsqualität auf der einen Seite und den hier gewählten Kriterien zur Erfassung von individuellen Auswirkungen auf der anderen Seite besteht. Für die Konstruktion der Entscheidungsmatrix werden somit die Konstrukte WISS (Wissen), VERT (Vertrauen), RISK (Risiko) und LOYA (Loyalität) als Zielkriterien für Auswirkungen eines Portals herangezogen.

6.3.2 Ergebnisse und Interpretation

Voraussetzung für einen sinnvollen Vergleich der Umfrageergebnisse zwischen den Portalen ist eine zumindest grobe Übereinstimmung der Eigenschaften der Nutzer der verschiedenen Portale. Dazu werden die Ergebnisse des demographischen Teils des Fragebogens je Portal betrachtet. Anschließend wird die Präferenzstruktur des Entscheidungsträgers bezüglich der Nutzermeinungen ermittelt und die Entscheidungsmatrix konstruiert.

Demographische Auswertung

Da es sich bei den demographischen Daten um Kontextinformationen handelt, die für die Bewertung selbst nicht herangezogen werden, wird auf eine aufwändige Imputation verzichtet und ein geringer Umfang an fehlenden Werten in Kauf genommen. Alle vorhandenen Daten (auch die aus den unvollständig ausgefüllten Fragebögen) werden dazu herangezogen.

Der Vergleich der abgebildeten Profile⁶ von Nutzergruppenzugehörigkeit (Abbildung 6.8), Nutzungshäufigkeit (Abbildung 6.9) und Nutzungsgründen (Abbildung 6.10) der Nutzer der verschiedenen Portale zeigt große Übereinstimmungen. Nur bei Portal P3 weicht die Zusammensetzung der Nutzer sowie des Schemas der Nutzungshäufigkeit deutlich von den anderen Portalen ab. Im Vergleich zu diesen gehörte die Mehrheit der Befragten von P3 nicht zur Gruppe der Lehrer und besuchte das Portal seltener.

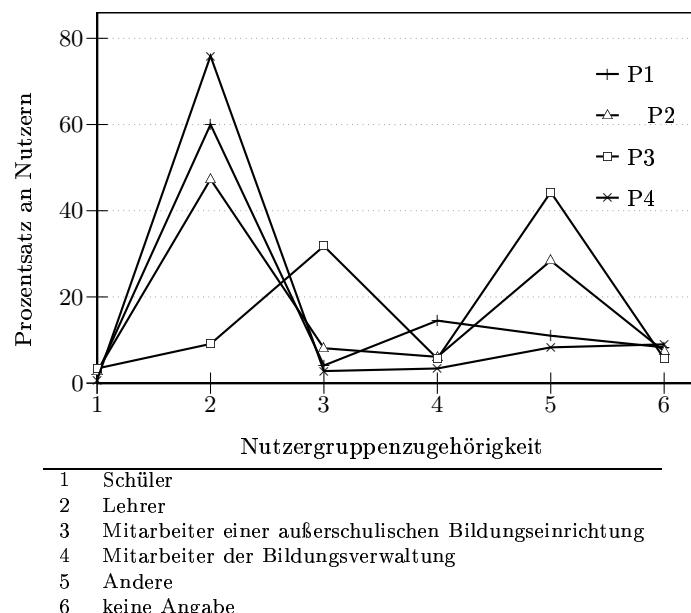


Abbildung 6.8: Prozentuale Verteilung der Nutzergruppen

Der Vergleich der Erfahrungen im Umgang mit dem Portal (Portalerfahrung) und im Umgang mit dem Web im Allgemeinen (Weberfahrung) zeigt, dass der Großteil aller Nutzer sehr erfahren ist (Tabelle 6.16).

Eine Clusteranalyse⁷ der Portale zu Nutzergruppenzugehörigkeit, Nutzungshäufigkeit, Portalerfahrung, Weberfahrung und Nutzungsgründen verdeutlicht nochmals die oben genannten Unterschiede. Die Portale P1, P2 und P4 fallen stärker

⁶Die Verbindungslienzen zwischen den Punkten dienen nur der besseren Vergleichbarkeit und stellen keine Kurven oder Geraden dar.

⁷Für die Clusteranalyse wurde der im Softwaretool WEKA (Witten u. Frank 2005) implementierte EM-Algorithmus verwendet. Dieser bestimmt die optimale Anzahl an Clustern und ermöglicht die Analyse kategorischer Daten.

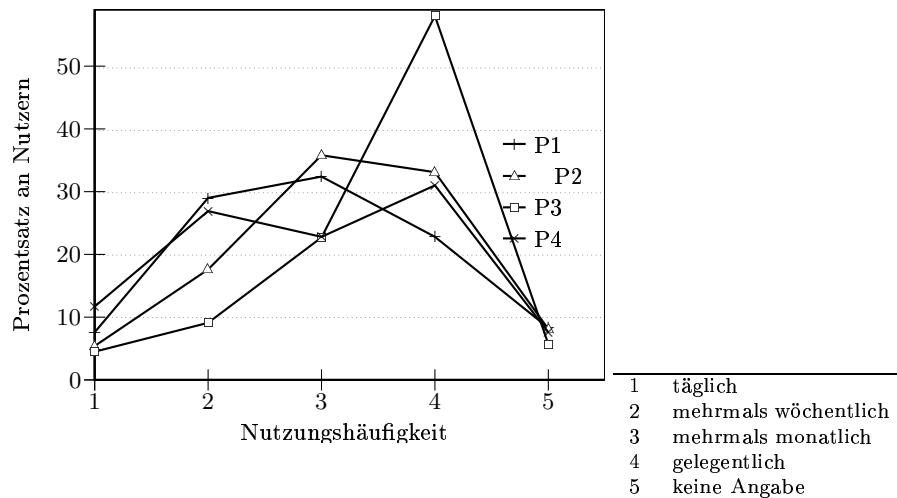


Abbildung 6.9: Prozentuale Verteilung der Nutzungshäufigkeiten

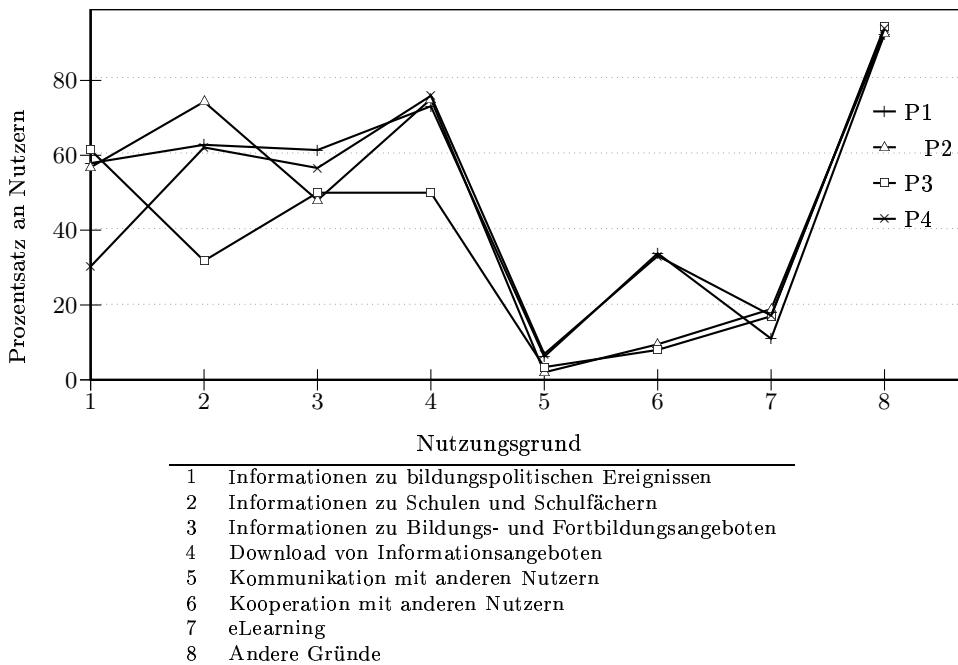


Abbildung 6.10: Prozentuale Verteilung der angegebenen Nutzungsgründe

in den Clustern zusammen (Cluster 1–3), als es bei Portal P3 der Fall ist (Cluster 0), siehe Abbildung 6.11.

Die vorhandenen Unterschiede könnten sich hier in einem bei den Nutzern von P3 im Vergleich zu den anderen drei Portalen abweichenden Bewertungsschema auswirken. Das ist bei der Interpretation der Ergebnisse zu berücksichtigen. Insgesamt kann jedoch aus den demographischen Daten schließen, dass die

Tabelle 6.16: Anteile der Befragten Nutzer mit mehr als 12 Monaten Portalerfahrung und mehr als fünf Jahren Weberfahrung

	Portal	Portalerfahrung	Weberfahrung
P1	71%	70%	
P2	66%	78%	
P3	74%	85%	
P4	75%	76%	

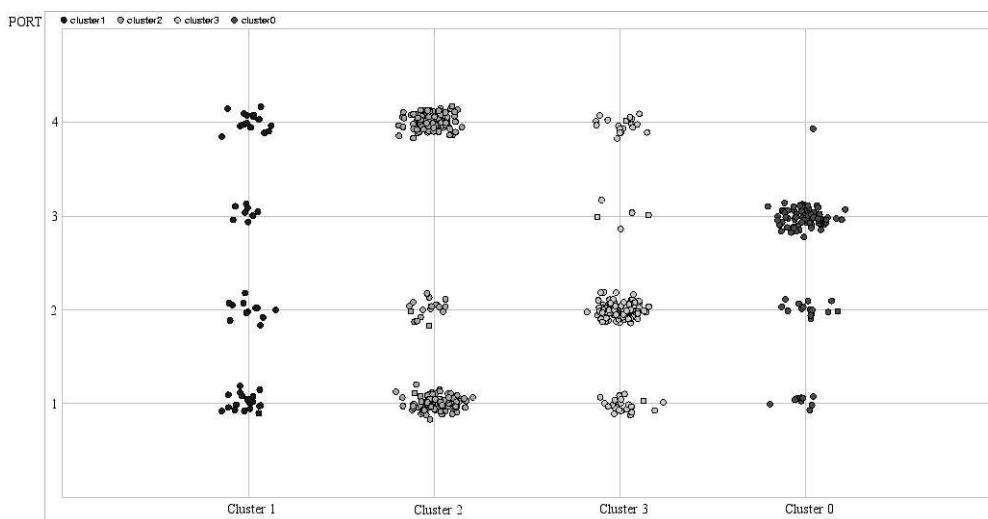


Abbildung 6.11: Clusterergebnis auf Basis der demographischen Daten

Nutzereigenschaften in großen Bereichen übereinstimmen und ein Vergleich der Umfrageergebnisse zwischen den Portalen daher sinnvoll ist.

Präferenzstruktur des Entscheidungsträgers

Für die Konstruktion der Entscheidungsmatrix sind zunächst die Gewichtungsfaktoren auf Basis der Präferenzvorstellungen des Entscheidungsträgers zu ermitteln. Dieses erfolgt analog zur beschriebenen Vorgehensweise in Abschnitt 6.2 mittels Ratingskalen. Der Entscheider wird gebeten, die Wichtigkeit der Kriterien zu den individuellen Auswirkungen anzugeben. Daraus werden dann unter Annahme einer Intervalskalierung der Ratingskala die Gewichtungsfaktoren abgeleitet und dem Entscheider zur Bestätigung vorgelegt (Tabelle 6.17). Die Ratings sind im Allgemeinen sehr hoch, was auf die Wichtigkeit dieser Kriterien aus Sicht des Entscheidungsträgers hinweist. Diese können daher mit Berechtigung als Erfolgsindikatoren bzw. Kriterien in der Entscheidungsmatrix eingesetzt werden. Ausnahme sind die Kriterien bzw. Items zur Erfassung der Auswirkungen auf die Wissensbildung, welche nur einen mittleren Stellenwert im Vergleich zu den anderen haben.

Tabelle 6.17: Präferenzen zu den individuellen Auswirkungen

Kriterium	Rating (1–5)	Gewichtung Item	Ratingssumme	Gewichtung Konstrukt
WISS1	3	0,60		
WISS2	2	0,40		
WISS			5	0,152
VERT1	5	0,56		
VERT2	4	0,44		
VERT			9	0,273
RISK1	5	0,56		
RISK2	4	0,44		
RISK			9	0,273
LOYA1	5	0,50		
LOYA2	5	0,50		
LOYA			10	0,303
Σ			33	~ 1

Entscheidungsmatrix

Zur Konstruktion der Entscheidungsmatrix sind für jedes Portal die Quartile ($\hat{Q}^{(1)}$, $\hat{Q}^{(2)}$, ..., $\hat{Q}^{(5)}$) aus den $m = 5$ imputierten Datentabellen zu ermitteln und in einem Mittelwert (\bar{Q}) zusammenzuführen. Die Werte für Portal P2 sind in Tabelle 6.18 aufgeführt ($\sqrt{\hat{B}}$ ist die Standardabweichung vom Mittelwert \bar{Q} , die sich aus den fünf Quartilwerten ergibt; siehe Anhang A).

Tabelle 6.18: Quartile zum Umfrageergebnis für P2

Item	1. Quartil							Median							3. Quartil							
	$\hat{Q}^{(1)}$	$\hat{Q}^{(2)}$	$\hat{Q}^{(3)}$	$\hat{Q}^{(4)}$	$\hat{Q}^{(5)}$	\bar{Q}	\sqrt{B}	$\hat{Q}^{(1)}$	$\hat{Q}^{(2)}$	$\hat{Q}^{(3)}$	$\hat{Q}^{(4)}$	$\hat{Q}^{(5)}$	\bar{Q}	\sqrt{B}	$\hat{Q}^{(1)}$	$\hat{Q}^{(2)}$	$\hat{Q}^{(3)}$	$\hat{Q}^{(4)}$	$\hat{Q}^{(5)}$	\bar{Q}	\sqrt{B}	
INFQ1	4	4	4	4	4	4	0	4	4	4	4	4	4	0	5	5	5	5	5	5	0	
INFQ2	4	4	4	4	4	4	0	4	4	4	4	4	4	0	5	5	5	5	5	5	0	
INFQ3	3	3	3	3	3	3	0	4	4	4	4	4	4	0	4	4	4	4	4	4	0	
INFQ4	3	3	3	3	3	3	0	4	4	4	4	4	4	0	4	4	4	4	4	4	0	
SYSQ1	3	3	3	3	3	3	0	4	4	4	4	4	4	0	4	4	4	4	4	4	0	
SYSQ2	4	4	4	4	4	4	0	4	4	4	4	4	4	0	5	5	5	5	5	5	0	
SYSQ3	3	3	3	3	3	3	0	4	4	4	4	4	4	0	5	4*	4	4	4*	4,2	0,4472	
SYSQ4	3	3	3*	3*	3	3	0	4	4	4	4	4	4	0	5	5	5	5	5	5	0	
SYSQ5	3	3	3	3	3	3	0	4	4	4	4	4	4	0	5	5	5	5	5	5	0	
SYSQ6	3	2*	3	3	3	3	2,8	0,4472	4	3	3	3	3	3,2	0,4472	4	4	4	4	4	4	0
WISS1	3	3	3	3	3	3	0	4	4	4	4	4	4	0	5	5	5	5	5	5	0	
WISS2	3	3	3	3	3	3	0	3	3	3	3	3	3	0	4	4	4	4	4	4	0	
VERT1	3	3	3	3	3	3	0	4	4	4	4	4	4	0	4	4	4	4	4	4	0	
VERT2	4	4	4	4	4	4	0	5	5	5	5	5	5	0	5	5	5	5	5	5	0	
RISK1	4	4	4	4	4	4	0	5	5	5	5	5	5	0	5	5	5	5	5	5	0	
RISK2	4	3	4	4	2*	3,4	0,8944	4	4	5	5	4	4	4,4	0,5477	5	5	5	5	5	5	0
LOYA1	3	3	3	3	3	3	0	4	4	4	4	4	4	0	5	5	5	5	5	5	0	
LOYA2	3	3	3	3	3	3	0	4	4	4	4	4	4	0	5	5	5	5	5	5	0	

* Unterquartil

Da hier von der Annahme ausgegangen wird (siehe Abschnitt 5.2.8), dass die auf dem Fragebogen genutzte Ratingskala ordinalskaliert ist, darf ein Quartil nur einen ganzzahligen Wert annehmen. Für den Fall, dass dieses genau zwischen zwei Elementen fällt, wird daher die in der geordneten Stichprobe zuerst auftretende Zahl (das Unterquartil) gewählt. Die entsprechenden Fälle für P2 sind in Tabelle 6.18 mit „*“ markiert. Um ganzzahlige Werte für die Mittelwerte \bar{Q} der Quartile aus den imputierten Datensätzen zu erhalten, werden diese bei Bedarf gerundet. Im Fall von P2 ist das nur für Item RISK2 beim 1. und 2. Quartil notwendig. Die Herleitung der Quartile für die anderen Portale erfolgt analog.

Die Umfrageresultate sind insgesamt sehr positiv. Die Mediane bewegen sich je nach Frage zwischen „4“ und „5“. Ausnahme ist Frage WISS2 („Die Informationen helfen mir beim Treffen von Entscheidungen“) mit einem Median von „3“. Da jedoch alle Portale für WISS2 diese Bewertung erhalten haben, ist diese im Vergleich also relativ gut.

Die schlechtere Bewertung bei WISS2 und die recht guten Bewertungen zur Informationsqualität (INFQ1-INFQ4) deuten darauf hin, dass die Nutzer das Portal eher als Informationsquelle, denn als System zur Entscheidungsunterstützung, sehen.

Da das Entscheidungsmodell zu den Nutzermeinungen für die Erfassung der Auswirkungen konstruiert ist, werden die Ergebnisse zu den INFQ- und SYSQ-Items nicht benötigt. Diese können aber als Kontextinformationen herangezogen werden, um die auf Basis der Konstrukte WISS, VERT, RISK und LOYA ermittelten individuellen Auswirkungen besser zu interpretieren.

Die Entscheidungsmatrix (Tabelle 6.19) wird wie in Beispiel in Tabelle 5.8 konstruiert und die Utility Scores werden mittels Gleichung 5.10 errechnet. Zur besseren Übersicht sind die Werte der Quartile in der hier abgebildeten Entscheidungsmatrix entsprechend dem innersten Teil von Gleichung 5.10 bereits aggregiert.

Tabelle 6.19: Entscheidungsmatrix zu den Nutzermeinungen

w_j	0,15 WISS		0,27 VERT		0,27 RISK		0,30 LOYA		$U_{\text{SAW}}^{\text{Nm.}}$
c_j	w_{jk}	c_{jk}	w_{jk}	c_{jk}	w_{jk}	c_{jk}	w_{jk}	c_{jk}	
P1	0,8000	0,6667	0,7333	0,9333	0,9333	0,8000	0,8000	0,8000	0,818
P2	0,8000	0,6667	0,7333	0,9333	0,9333	0,8000	0,8000	0,8000	0,818
P3	0,7333	0,6000	0,7333	0,9333	0,9333	0,8000	0,7333	0,8000	0,798
P4	0,6667	0,6000	0,6667	0,9333	0,9333	0,8000	0,8000	0,8000	0,792

Die Werte für die Gewichtungsfaktoren sind gerundet.

P2 erzielt auf Basis der gewählten Kriterien und Gewichte zusammen mit P1 das beste Ergebnis. Daraus lässt sich schließen, dass das Portal als informativ, vertrauenswürdig und risikoarm eingestuft und diesem eine hohe Loyalität entgegengebracht wurde.

Wie aus der Matrix ersichtlich, erzielt P2 für fast alle Kriterien im Vergleich gute bis sehr gute Bewertungen. Die Präferenzvorstellungen können durch die für alle Kriterien erzielten hohen Werte als erfüllt angesehen werden.

Insgesamt haben alle Portale relativ gute Bewertungen erhalten, mit nur geringen Unterschieden. Dieses deutet auf eine von den Nutzern als allgemein gut empfundene Dienstleistungsqualität bei allen Portalen hin.

Im Vergleich weist P2 aufgrund der gewählten Kriterien und Gewichte bezüglich der individuellen Auswirkungen der Portalnutzung keine Schwachstellen auf.

6.3.3 Diskussion

Ziel der Analyse der Nutzermeinungen ist es, Auswirkungen der Portalnutzung aus der Perspektive der individuellen Wahrnehmung der Nutzer zu erfassen. Betrachtet wird der Einfluss auf (1) die Wissensbildung, (2) die Ausbildung von Vertrauen gegenüber dem Portal bzw. dem Portalbetreiber, (3) das empfundene Risiko der Portalnutzung sowie (4) die dem Portal entgegengebrachte Loyalität. Die zugrunde liegende Annahme entsprechend Proposition 2 ist hierbei, dass diese Variablen durch die Qualität von Portalsystem und bereitgestellten Informationen beeinflusst werden und daher als Auswirkungen der Portalnutzung betrachtet werden können. Das Ergebnis der empirischen Analyse zeigt, dass eine Abhängigkeit zwischen Portalsystem- und Informationsqualität auf der einen und den hier betrachteten Auswirkungen auf der anderen Seite vorliegt. Dieses weist auf die Gültigkeit von Proposition 2 hin.

Bezüglich der verwendeten Bewertungskriterien für Auswirkungen wird hier kein Anspruch auf Vollständigkeit erhoben. Gegenstand von Folgeuntersuchungen könnte zum einen die Bestätigung der hier Erfassten Zusammenhänge sein und zum anderen die Erforschung weiterer wichtiger Variablen individueller Auswirkungen, um das Instrument zu verbessern.

Zur Modellvalidierung wird der PLS-Ansatz verwendet. Dieser hat im Vergleich zu anderen Ansätzen (z. B. LISREL) weniger restriktive Voraussetzungen bezüglich der Verteilung der Daten (bzw. von der Normalverteilung abweichenden Daten) sowie der notwendigen Mindestanzahl an Fällen (siehe Fornell u. Bookstein 1982). Außerdem eignet sich PLS gut für explorative Analysen, welche hier in Zusammenhang mit den gewählten Kriterien zur Erfassung der Auswirkungen vorliegt. Alternativ könnte in zukünftigen Untersuchungen ein anderer Ansatz gewählt werden, um zu ermitteln, inwieweit das hier vorgeschlagene Modell auch restriktiveren Annahmen Stand hält.

Voraussetzung für einen sinnvollen Vergleich der Auswertungsergebnisse der Portale ist eine zumindest grobe Übereinstimmung der Eigenschaften ihrer Nutzer. Nur so kann man von einem ähnlichen Bewertungsschema bei den Nutzern der verschiedenen Portale ausgehen. Die Ergebnisse des demographischen Teils des Fragebogens deuten diesbezüglich auf eine ausreichende Übereinstimmung hin. Die gesammelten Daten liefern allerdings zu wenige Informationen, um über die Stärke dieser Übereinstimmung genaue Aussagen ableiten zu können. Man muss hier

davon ausgehen, dass die Nutzergruppen der verschiedenen Portale bis zu einem gewissen Grad in ihren Eigenschaften doch von einander abweichen und diese Unterschiede sich auch möglicherweise auf das Antwortverhalten und damit auf die Umfrageergebnisse auswirken. Wie groß dieser Einfluss ist, wäre eine Fragestellung für zukünftige Untersuchungen.

Die Auswertungsergebnisse zu den Portalen differieren nur in geringem Maße voneinander. Das kann darauf zurückzuführen sein, dass bei allen untersuchten Portalen ein hohes Qualitätsniveau vorliegt. Es kann aber auch der Fall sein, dass die Anzahl der Stufen (1–5) der für den Fragebogen gewählten Ratingskala für eine ausreichende Differenzierbarkeit zu gering ist. In Folgeuntersuchungen könnte ermittelt werden, ob beispielsweise eine 7- oder 10-Stufen-Skala zu differenzierteren Ergebnissen führen würde.

Ein häufiges Problem bei schriftlichen Umfragen ist die Nicht-Beantwortung von Fragen. Die Standardvorgehensweise bei einer Auswertung ist es, unvollständige Fragebögen zu ignorieren. Dadurch verliert man im besten Fall nur zusätzliche Informationen. Im schlimmsten Fall erhält man so Ergebnisse mit systematisch behafteten Fehlern. Die Ergänzung fehlender Werte durch Multiple Imputation führt in der Regel zu valideren Ergebnissen und sollte daher immer bevorzugt werden. Da die Qualität der Imputation vom gewählten Imputationsmodell abhängt, sollte dieses unter Berücksichtigung der zugrunde liegenden Daten gewählt werden. Die hier gewählte Imputationsmethode basiert auf logistischer Regression und ist für kategoriale Daten gut geeignet. Die Methode setzt eine multivariate Normalverteilung der Daten voraus. Obwohl die erhobenen Daten bzw. die Stichprobe nicht für alle Items eine entsprechende Verteilung vorweisen, führt diese Imputationsmethode bei einem mäßigen Anteil fehlender Werte dennoch zu robusten Auswertungsergebnissen. Die Anteile fehlender Werte bei den hier betrachteten Items bzw. Variablen liegen unter 35% und sind somit noch in einem vertretbaren Rahmen. Die imputierten Werte sollten daher eine relativ gute Approximation der wahren Werte darstellen. Da jedoch die wahre Verteilung der fehlenden Werte nicht bekannt ist, bleibt ein Rest an Unsicherheit.

6.4 Bewertung Aufwendungen

Die Vorgänge zur Pflege der Inhalte werden in den betrachteten Portalsystemen nicht festgehalten und können daher nicht objektiv erfasst werden. Die Zahlen werden mittels Befragung (Interview) der Portalbetreiber ermittelt.

6.4.1 Datenaufbereitung

Jeder Portalbetreiber wird gebeten, die durchschnittliche Anzahl wöchentlicher Contenterstellungs-, -update und -löschvorgänge zu schätzen. Anhand der ermittelten Anzahl an Sessions aus der Logdatenauswertung werden so die Contentänderungsverhältnisse für Erstellen ($cav(create)$), Aktualisieren ($cav(update)$) und Löschen ($cav(delete)$) je Nutzersession errechnet (Tabelle 6.20).

Tabelle 6.20: Contentänderungsverhältnisse pro Session (Werte in Klammern) auf Basis der ermittelten wöchentlichen Anzahl an Sessions und der geschätzten wöchentlichen Contentänderungsvorgänge durch den Portalbetreiber

Portal	Sessions	Erstellvorgänge	Updatevorgänge	Löschvorgänge
P1	15626	20 (0,001280)	2,5 (0,000160)	8 (0,000512)
P2	88383,25	50 (0,000566)	5 (0,000057)	2 (0,000023)
P3	162152	80 (0,000493)	30 (0,000185)	12,5 (0,000077)
P4	26530,5	3 (0,000113)	1 (0,000038)	0,5 (0,000019)

6.4.2 Ergebnisse und Interpretation

Die Präferenzstruktur des Entscheidungsträgers bezüglich der Aufwendungen wird ermittelt und die Entscheidungsmatrix konstruiert.

Präferenzstruktur des Entscheidungsträgers

Die Herleitung der Gewichtungsfaktoren auf Basis der Präferenzvorstellungen des Entscheiders erfolgt analog zu Abschnitt 6.2 mittels Ratingskala. Der Entscheider wird gebeten, die Wichtigkeit eines jeweiligen Contentänderungsvorgangs anzugeben. Daraus werden dann unter Annahme einer Intervalskalierung der Ratingskala die Gewichtungsfaktoren abgeleitet (Tabelle 6.21) und dem Entscheider zur Überprüfung vorgelegt. Die Ratings sind alle sehr hoch, was auf die Wichtigkeit dieser Kriterien aus Sicht des Entscheidungsträgers hinweist. Diese können daher mit Berechtigung als Erfolgsindikatoren bzw. Zielkriterien in der Entscheidungsmatrix eingesetzt werden.

Tabelle 6.21: Präferenzen zu den Aufwendungen

Vorgang	Rating (1–5)	Gewichtung
Erstellen (create)	5	0,3333
Aktualisieren (update)	5	0,3333
Löschen (delete)	5	0,3333
\sum	15	~ 1

Entscheidungsmatrix

Die partiellen Nutzenwerte für die Entscheidungsmatrix werden mittels der in Gleichung 5.11 angegebenen Nutzenfunktion hergeleitet. Je höher ein Nutzenwert, desto besser ist das Contentänderungsverhältnis pro Nutzersession. Die resultierende Entscheidungsmatrix zeigt, dass P2 auf Basis der gewählten Kriterien und Gewichte das zweitbeste Ergebnis erzielt (Tabelle 6.22). Der Contentpflegeaufwand bewegt sich somit, gemessen an der Zahl der Besuchersessions, in einem mittleren bis guten Bereich im Vergleich zu den anderen Portalen. Die Präferenzvorstellun-

gen können durch die für alle Kriterien erzielten, hohen Werte als erfüllt angesehen werden.

Tabelle 6.22: Entscheidungsmatrix zu den Aufwendungen

w_j	0,3333	0,3333	0,3333	$U_{\text{Aufwendungen}}^{\text{SAW}}$
c_j	Erstellen	Aktualisieren	Löschen	
P4	0,9887	0,9962	0,9981	0,994
P2	0,9434	0,9943	0,9977	0,979
P3	0,9507	0,9815	0,9923	0,975
P1	0,8720	0,9840	0,9488	0,934

Die Werte für die Gewichtungsfaktoren sind gerundet.

Aus der Umfrage (Abschnitt 6.3) geht hervor, dass die Mehrheit der Nutzer das Portal mehrmals monatlich bzw. gelegentlich besucht (Abbildung 6.9). Da bei P2 durchschnittlich recht viele Contenterstellungsvorgänge pro Woche vorgenommen werden, kann man davon ausgehen, dass in der Mehrzahl der Fälle neue Inhalte bei den wiederkehrenden Besuchern zur Verfügung stehen. Darauf deuten ebenfalls die im Allgemeinen guten Bewertungen der Nutzer zur Informationsqualität (INFQ1–INFQ4 in Tabelle 6.18).

6.4.3 Diskussion

Ziel der Analyse der Aufwendungen ist eine Effizienzbetrachtung der Contentpflegevorgänge in Bezug zur Anzahl an realisierten Besuchersessions. Als Kriterien werden die Contentänderungsverhältnisse für das Erstellen, Aktualisieren und Löschen herangezogen. Dahinter steht die Überlegung (siehe Abschnitt 5.3.3), dass die Pflege von Inhalten am effizientesten ist, wenn die Frequenz von Änderungen der Inhalte der Besuchsfrequenz wiederkehrender Nutzer entspricht.

Das Ergebnis zeigt, dass aufgrund der hier erfassten Daten eine ansatzweise Effizienzbetrachtung in Bezug auf die erzielten Effekte und im Vergleich zu den anderen Portalen möglich ist. Dieses muss hier allerdings eher als erster Hinweis denn als präzise Aussage gewertet werden, da (1) die zugrunde liegenden Daten für Contentänderungsvorgänge auf Basis der möglicherweise nicht exakten Schätzungen der Portalbetreiber ermittelt werden und (2) die untersuchten Logdaten zur Ermittlung der durchschnittlichen Anzahl an Sessions einen Zeitraum von maximal vier Wochen abdecken. Sowohl ein Mechanismus zur exakten Erfassung der Contentänderungsvorgänge als auch der wiederkehrenden Besucher sowie die Auswertung von Logdaten aus einem längeren Zeitraum (z. B. ein Jahr) würden zu präziseren Ergebnissen führen und damit validere Aussagen ermöglichen. Eine exakte Berechnung der mindestens notwendigen Anzahl an wöchentlichen Contentpflegevorgängen könnte dann außerdem bei der Einschätzung helfen, ob Änderungen in ausreichendem Maße vorgenommen werden. Ein entsprechender Mechanismus war bei den untersuchten Portalen nicht vorhanden und, wie in Abschnitt 6.2 bereits er-

läutert, war im Rahmen dieser Arbeit eine Auswertung größerer Logdatenmengen aufgrund des hohen Aufwands nicht möglich. In einer zukünftigen Studie könnte dieser Zusammenhang mittels einer exakten Erfassung von Contentpflegevorgängen und wiederkehrenden Besuchern empirisch untersucht werden.

Die in Proposition 3 getroffene Annahme, dass Portalpflegevorgänge sich umso bezahlter machen, je mehr Nutzer diese wahrnehmen, wird hier nicht empirisch validiert. Die vorhandenen Auswertungsergebnisse deuten zumindest daraufhin, dass die Contentpflege sich auf die empfundene Informationsqualität auswirkt und damit eine Mindestanzahl an Contentpflegevorgängen für einen erfolgreichen Portalbetrieb notwendig ist. Beispielsweise haben die Portalbetreiber von P1 und P2 zwar einen relativ hohen Contentpflegeaufwand im Vergleich zu P4 (Tabelle 6.22), dafür aber auch bessere Bewertungen bezüglich der Nutzermeinungen (Tabelle 6.19).

Insgesamt können die hier ermittelten Ergebnisse auf Grundlage der Komponente zur Erfassung der Aufwendungen nur als erster Hinweis auf eine Effizienzbe trachtung angesehen werden. Wie oben beschrieben, besteht noch deutlich Potential für zukünftige Verbesserungen.

6.5 Gesamtbewertung

In der Gesamtbewertung werden die Ergebnisse der drei in den vorangegangenen Abschnitten untersuchten Bereiche zusammengeführt.

6.5.1 Ergebnisse und Interpretation

Die Präferenzstruktur des Entscheidungsträgers bezüglich der Gesamtbewertung wird ermittelt und die Entscheidungsmatrix konstruiert.

Präferenzstruktur des Entscheidungsträgers

Die Herleitung der Gewichtungsfaktoren auf Basis der Präferenzvorstellungen des Entscheiders erfolgt analog zu Abschnitt 6.2 mittels Ratingskala. Der Entscheider wird gebeten, die Wichtigkeit jeder der drei hier betrachteten Bereiche, die die drei Komponenten der Gesamtbewertung darstellen, für die Erfolgsbewertung anzugeben. Daraus werden dann unter Annahme einer Intervalskalierung der Ratingskala die Gewichtungsfaktoren abgeleitet (Tabelle 6.23) und dem Entscheider zur Überprüfung vorgelegt.

Das Rating zur Komponente „Nutzerverhalten“ ist am höchsten, d. h. der Betreiber von Portal P2 sieht im Verhalten der Nutzer den wichtigsten Erfolgsindikator. Als etwas weniger wichtige Faktoren folgen gleichauf die Nutzermeinungen und die Aufwendungen. Die Einstufungen dieser zwei Komponenten zeigen, dass diese aus Sicht des Entscheidungsträgers als Kriterien für die Gesamtbewertung des Erfolgs von P2 zwar eine nicht unwichtige Rolle spielen, es aber noch andere wichtigere (hier nicht erfasste) Kriterien zu geben scheint.

Tabelle 6.23: Präferenzen zur Gesamtbewertung

Kriterium	Rating (1–5)	Gewichtung
Nutzerverhalten	4	0,4
Nutzermeinungen	3	0,3
Aufwendungen	3	0,3
Σ	10	1

Entscheidungsmatrix

Die Entscheidungsmatrix wird aus den drei betrachteten Komponenten konstruiert und der Gesamt-Utility Score wird entsprechend Definition 15 (Abschnitt 5.3.4) berechnet. Eine gesonderte Herleitung partieller Nutzenwerte ist nicht mehr notwendig, da die Utility-Scores der einzelnen Komponenten bereits Nutzenwerte darstellen. Der Erfolg bzw. das Erfolgsniveau eines Portals wird entsprechend Definition 16 (Abschnitt 5.3.4) interpretiert.

Die resultierende Entscheidungsmatrix zeigt, dass P2 auf Basis der betrachteten Komponenten und der gewählten Gewichte das zweitbeste Gesamtergebnis erzielt (Tabelle 6.24). Bei diesem Portal liegt ein hohes Erfolgsniveau in allen betrachteten Bereichen relativ zu den Ergebnissen der anderen Portale vor. Schwerwiegende Schwachstellen sind somit aufgrund der vorhandenen Daten nicht erkennbar.

Dennoch besteht möglicherweise Verbesserungspotential in dem für den Betreiber von P2 als am wichtigsten erachteten Bereich, dem Nutzerverhalten. Der Vergleich mit dem Ergebnis von Portal P1 ergibt einen deutlichen Unterschied. Wie in Abschnitt 6.2 detailliert beschrieben, könnte bei P2 möglicherweise ein besseres Ergebnis durch entsprechende, an die Umsetzung in Portal P1 angelehnte, Änderungen in der Portalstruktur sowie im Aufbau von Navigations- und Content-Seiten erzielt werden.

Insgesamt gibt es – bis auf das Kriterium „Nutzerverhalten“ – keine extremen Unterschiede in den Ergebniswerten zwischen den Portalen. Das deutet auf ein in großen Teilen ähnliches Qualitätsniveau der Bereitstellung von Diensten und Informationen auf den betrachteten Portalen hin.

Tabelle 6.24: Entscheidungsmatrix zur Gesamtbewertung

w_j c_j	0,4 Nutzerverhalten	0,3 Nutzermeinungen	0,3 Aufwendungen	$U_{\text{Gesamt}}^{\text{SAW}}$
P1	0,2941	0,8182	0,9349	0,644
P2	0,1465	0,8182	0,9785	0,598
P4	0,1325	0,7919	0,9943	0,589
P3	0,1108	0,7980	0,9748	0,576

6.5.2 Diskussion

In der Gesamtbewertung werden die Ergebnisse zu den einzelnen Komponenten des hier entwickelten Entscheidungsmodells zusammengeführt und zu einem Gesamturteil verdichtet. Die zugrunde liegende Annahme ist, dass das Nutzerverhalten, die Nutzermeinungen sowie die Aufwendungen zur Pflege von Content wichtige und ausreichende Indikatoren darstellen, um aus deren Betrachtung sinnvolle Aussagen zu Erfolg und Verbesserungspotential eines Portals ableiten zu können. Zwei wichtige Argumente für die Gültigkeit dieser Kriterien als Erfolgsindikatoren sind (1) deren besonderer Stellenwert in der Literatur, insbesondere das Nutzerverhalten und die Nutzermeinungen (siehe Abschnitt 3.3), und (2) die hier zu diesen Kriterien abgegebenen hohen Präferenzratings seitens des Entscheidungsträgers.

Die sich daran anschließende Frage ist die der Operationalisierung dieser Kriterien bzw. Kategorien. Auch hier kann zur Begründung der gewählten Operationalisierungskriterien auf die in der Literatur bestehenden und akzeptierten Vorschläge (siehe Abschnitt 3.2.8) sowie auf die hier abgegebenen hohen Präferenzratings seitens des Entscheidungsträgers verwiesen werden.

Die Erfolgsbetrachtung eines Portals anhand des hier vorgeschlagenen Entscheidungsmodells erfolgt ausschließlich auf Basis der verwendeten quantitativen Kriterien. Dadurch wird sichergestellt, dass alle betrachteten Portale auf der gleichen Bewertungsgrundlage analysiert und die Ergebnisse zweier Portale miteinander verglichen werden können. Dieser festgelegte Satz an Bewertungskriterien setzt voraus, dass die Portale große Ähnlichkeiten in ihren Eigenschaften aufweisen. Nur so kann eine sinnvolle Bewertung bzw. ein sinnvoller Vergleich gewährleistet werden. Dennoch können für ein Portal gewisse erfolgsrelevante Faktoren eine Rolle spielen, die nicht in dem Modell als Kriterien erfasst sind. Das gilt insbesondere für Charakteristiken eines Portals, die bei keinem anderen der zum Vergleich herangezogenen existieren. In solch einem Fall können diese nicht berücksichtigt werden und die Bewertung ist damit für dieses Portal nicht vollständig oder sogar falsch.

Ein ähnliches Problem gibt es mit qualitativen Kriterien. Diese können bestenfalls als Kontextinformationen im Rahmen der quantitativen Ergebnisauswertung herangezogen werden. Einen direkten Einfluss haben entsprechende und unter Umständen wichtige Faktoren hier nicht.

Für die anhand der verwendeten Kriterien erfassten Ergebniswerte gibt es keine „Standards“, die zum Zweck der Erfolgsbeurteilung herangezogen werden könnten. Trotz großer Übereinstimmungen in den Eigenschaften zweier Portale, können deren Portalbetreiber unterschiedliche Schwerpunkte in der Bereitstellung von Diensten und Informationen und damit unterschiedliche Zielsetzungen und Ansichten bezüglich eines erfolgreichen Portalbetriebs haben. Die Einbindung der Präferenzstruktur des Entscheidungsträgers in das Bewertungsmodell ermöglicht daher eine individualisierte Betrachtung von Erfolg. Inwieweit von Erfolg gesprochen werden kann, hängt damit maßgeblich von der Sichtweise des jeweiligen Entscheidungsträgers bzw. Stakeholders ab.

6.6 Schlussfolgerungen für den Portalbetrieb

Die Ergebnisse der Untersuchung werden dem Betreiber von Portal P2 mit der Bitte um Stellungnahme mitgeteilt. Von Interesse ist, inwiefern die Untersuchung hilfreich für eine Einstufung des Erfolgsniveaus, für das Aufdecken von Schwachstellen und für die Entwicklung von Verbesserungskonzepten war.

Nach Ansicht des Betreibers fußt das Analyseinstrumentarium auf einer soliden Grundlage. Die Resultate werden daher ernst genommen und als interessant und hilfreich eingestuft. Zusätzlich wird positiv bemerkt, dass erstmalig durch den quantitativen Vergleich mit anderen Bildungsservern ein Anhaltspunkt zum eigenen Leistungs niveau gegeben wurde.

Die sich aus dem Vergleich mit den anderen Portalen ergebende Positionierung von P2 ist für den Betreiber ein Hinweis auf die relativ gute Qualität der geleisteten Arbeit bei der Bereitstellung von Inhalten.

Die Ergebnisse zu den Nutzermeinungen zeigen dem Betreiber, dass die bereitgestellten Informationen aus Sicht der Nutzer eine hohe Qualität hatten und dem Portal ein hohes Vertrauen entgegengebracht wurde.

Die Ergebnisse zum betrachteten Arbeitsaufwand liefern dem Betreiber einen ersten Hinweis darauf, dass die investierte Arbeit im Vergleich zur Größe des Portals, der Anzahl an Nutzern und zu den anderen Portalen nicht aus dem Rahmen fiel.

Den Präferenzen des Betreibers entsprechend, sind die Ergebnisse zum Nutzerverhalten von besonderem Interesse:

Mit Ausnahme von Zugriffen auf Seiten des Typs M sind die Effektivitätswerte für alle anderen Seitentypen sowie herunterladbare Dokumente relativ gering (siehe Tabelle 6.9). Der Betreiber von P2 schließt daraus, dass auf höheren Hierarchieebenen klarer dargestellt werden müsse, welche Inhalte auf den jeweils unteren Ebenen vorzufinden sind. Zugriffe könnten so zielgerichteter erfolgen und es gäbe weniger „ineffektive“ Zugriffe. Des Weiteren sieht der Betreiber den Bedarf, diese stärker zielgerichtete Navigation mittels Ersetzen reiner Navigations-Seiten durch Content-Navigations-Seiten des Typs MNINE in vielen Fällen zu unterstützen. Auf diese Weise könnten einem Nutzer schon frühzeitig während des Navigierens zur erwünschten Zielseite Informationen über den zu erwartenden Inhalt gegeben werden. Ein Nutzer könnte dann früher entscheiden, ob der Inhalt dem gesuchten entspricht oder ob an anderer Stelle weitergesucht werden soll.

Aus dem geringen Effektivitätswert bei Zugriffen auf herunterladbare Dokumente, schließt der Betreiber von P2, dass hier analog zu obigem eine klarere Darstellung der zu erwartenden Inhalte auf den höheren Hierarchieebenen notwendig sei. Da über das Portal in vielen Fällen auch zu Dokumenten auf anderen Quellen bzw. Websites verlinkt wurde, sah der Betreiber ein, dass der Anteil an direkt vom Portal herunterladbaren Dokumenten eventuell nicht sonderlich erhöht werden könne. Ein Überdenken der Präferenzen wäre somit hier eventuell nötig.

Die Effizienz von Zugriffen hatte für den Betreiber von P2 eine geringere Präferenz als die Effektivität (siehe Abbildung 6.4). Dennoch sah dieser aufgrund der

Ergebnisse die Notwendigkeit, reine Navigations-Seiten mit nur sehr wenigen Links (z. B. weniger als fünf) durch entsprechende Umstrukturierungen zu MNINE-Seiten umzugestalten oder aber durch andere Navigations-Seiten vollständig zu ersetzen. Auf diese Weise sollte eine unnötige Hierarchietiefe des Portals vermieden und das Navigieren für die Nutzer komfortabler werden.

6.7 Zusammenfassung

In diesem Kapitel wurde das in Kapitel 5 entwickelte Bewertungsmodell im Rahmen einer Fallstudie von vier Bildungsservern angewendet. Die Analyse erfolgte aus der Perspektive und unter Einbezug der Präferenzstruktur des Betreibers von Portal P2.

Die Ergebnisse weisen Performanceunterschiede zwischen den Portalen auf, woraus Schwächen und Stärken des Portals P2 ableitbar sind. Im Wesentlichen sind die Unterschiede aber eher gering, was auf ein zu großen Teilen ähnliches Qualitätsniveau bei den Portalen hinweist.

Die empirischen Ergebnisse zur Analyse des Nutzerverhaltens lassen Annahmen, dass Proposition 1 Gültigkeit besitzt. Demnach können aus der Effektivität und Effizienz von Zugriffen Schwachstellen in der Menü- und Portalstruktur identifiziert werden sowie Aussagen zum entgegengesetzten Interesse für einen bestimmten Seitentyp bzw. für herunterladbare Dokumente gemacht werden.

Die empirischen Ergebnisse zur Analyse der Nutzermeinungen zeigen, dass ein signifikanter Zusammenhang zwischen Qualitätsmerkmalen eines Portals und den individuell wahrgenommenen Auswirkungen besteht. Damit stützen die Ergebnisse Proposition 2.

Die Betrachtung der Aufwendungen liefert erste Hinweise auf einen Zusammenhang zwischen Contentpflegeaufwand und empfundener Informationsqualität und damit indirekt dem Portalerfolg. Die in Proposition 3 darüber hinaus gehende Annahme, dass sich Contentänderungen umso bezahlter machen je mehr Nutzer diese wahrnehmen, kann aufgrund der hier durchgeföhrten Untersuchung allerdings nicht direkt bestätigt werden. Eine weiterführende empirische Analyse wäre zur Klärung dieser Frage hilfreich.

Das Feedback des Betreibers von Portal P2 zeigt, dass die basierend auf dem vorgeschlagenen Bewertungsmodell gewonnenen Erkenntnisse auch von praktischem Nutzen für die Erfolgsbewertung und Verbesserung sind.

Kapitel 7

Zusammenfassung und Ausblick

Vorliegendes Kapitel stellt den in dieser Arbeit vorgeschlagenen Ansatz zusammenfassend dar, hebt die aus der Fallstudie gezogenen Schlussfolgerungen hervor und gibt abschließend einen Ausblick zu möglichen zukünftigen Fragestellungen, die sich aus den Ergebnissen der Arbeit ergeben.

7.1 Überblick

Der Einsatz von Portalen zur Repräsentation der Online-Präsenz von Organisationen hat in den letzten Jahren stetig zugenommen. Die speziellen Charakteristiken von Portalen machen diese zu geeigneten Informations-, Kommunikations- und Interaktionsplattformen zwischen den Mitgliedern einer Organisation, aber auch zwischen der Organisation und Nicht-Mitgliedern (z. B. Besuchern, Kunden, Lieferanten).

Auch nicht-kommerzielle Organisationen setzen in zunehmendem Maße Portale ein. Im Gegensatz zu kommerziellen Portalen, werden diese aber nicht mit dem Ziel der Profitgenerierung betrieben. Dennoch unterliegen dem Betrieb in der Regel gewisse Zielsetzungen, wie beispielsweise die Bereitstellung informativer Inhalte und hilfreicher Funktionen für die Nutzer.

Der Gebrauch nicht-kommerzieller Portale ist zwar für den Nutzer kostenfrei, der Betrieb hingegen verursacht Kosten. Die Ausgaben zu rechtfertigen und Möglichkeiten aufzudecken, dass Angebot eventuell zu verbessern, sind zusammen mit einem in der Regel beschränkten Budget daher wichtige Gründe, aus denen sich für nicht-kommerzielle Organisationen der Bedarf an einer Erfolgsbewertung ergibt. Das Hauptproblem ist hierbei, aussagekräftige Kriterien zu finden. Indikatoren wie Umsatz oder Gewinn, die im kommerziellen Bereich gebräuchlich sind, können nicht genutzt werden. Für den nicht-kommerziellen Bereich sind hier keine derartigen „Standardkriterien“ bekannt.

In der vorliegenden Arbeit wird ein Bewertungsansatz vorgeschlagen, welcher auf der Betrachtung von Nutzerverhalten, Nutzermeinungen und Aufwendungen zur Pflege von Content als Bewertungskategorien basiert. Die Interpretation der Ergebnisse erfolgt mittels Benchmarking. Dieser Vergleich bildet die Basis bzw. den Maßstab zur Bewertung.

Nutzerverhalten und Nutzermeinungen sind die wohl am häufigsten betrachteten Bereiche in der IS-Evaluationsforschung. Aufwendungen bzw. Kosten sind eine weitere häufig genutzte Bewertungskategorie. Die Operationalisierung dieser Kategorien basiert weitgehend auf in der Literatur verwendeten und akzeptierten Kriterien. Im Fokus des Bewertungsansatzes liegt die Betrachtung der Qualität der bereitgestellten Informationen und der Qualität der Informationsbereitstellung.

Gegenstand der Analyse des Nutzerverhaltens ist das Navigationsverhalten der Nutzer bei der Suche und Anfrage von Informationsinhalten. Die Untersuchung der Nutzermeinungen beinhaltet die individuell wahrgenommenen Effekte der Portalnutzung. Die Analyse der Aufwendungen betrachtet Vorgänge zur Pflege von Inhalten. Die Ergebnisse aus diesen drei Bereichen werden in einem Gesamtbewertungsmodell integriert und eine Gesamterfolgsbewertung wird durchgeführt.

Da es keine Standards gibt, kann das aus den Ergebnissen abgeleitete Erfolgsniveau von den jeweiligen Stakeholdern unterschiedlich eingeschätzt werden. Um diesen verschiedenen Sichtweisen Rechnung zu tragen, werden hier Stakeholderpräferenzen in die Bewertung einbezogen.

Die Umsetzung des Bewertungsansatzes erfolgt mittels Multikriterieller Ent-

scheidungsanalyse. Entsprechend wird ein multiattributives Entscheidungsmodell mit den Komponenten „Nutzerverhalten“, „Nutzermeinungen“ und „Aufwendungen“ entwickelt und angewendet.

7.2 Schlussfolgerungen

Das Bewertungsmodell wird im Rahmen einer Fallstudie von vier deutschen Bildungsserver-Portalen eingesetzt. Die Auswertung der empirischen Ergebnisse führt zu den in den folgenden Abschnitten zusammenfassend dargestellten Schlussfolgerungen.

7.2.1 Nutzerverhalten

Die Ergebnisse deuten darauf hin, dass Proposition 1 Gültigkeit besitzt, d. h. anhand der hier verwendeten Kriterien der Effektivität und Effizienz von Zugriffen lassen sich Rückschlüsse auf die Qualität von Portalstruktur und -inhalten ziehen. Eine direkte empirische Validierung der Proposition wird in dieser Arbeit jedoch nicht vorgenommen.

Der Vergleich mit dem Nutzerverhalten auf anderen Portalen ermöglicht eine recht gute Analyse, wie die Strukturierung von Portal und Seiteninhalten auf das Nutzerverhalten Einfluss nimmt. Die Identifizierung von good oder best practices ist hier möglich.

Der Fokus einer Untersuchung kann sich auf das ganze Portal oder aber auf Teilbereiche richten. Zu berücksichtigen ist, dass der betrachtete Bereich homogen in seiner Strukturierung sein sollte. Beispielsweise dürften sich in diesem Bereich die durchschnittlichen Pfadlängen zu Seiten des gleichen Typs (z. B. Typ M) nicht extrem von einander unterscheiden, da sonst die ermittelten Ergebniswerte eher eine Art „Mittelwert“ darstellten und damit ihre Aussagekraft abgeschwächt würde. Die untersuchten Bildungsserver sind in ihrer Gesamtstruktur untereinander weitgehend homogen.

Für eine regelmäßige Anwendung in der Praxis ist die Abstraktion der Requests durch Klassifikation der entsprechenden Seiten mittels Auslesen des Verhältnisses von Hyperlinks zu Text unpraktikabel und fehlerbehaftet. Außerdem ist der Bearbeitungsaufwand recht groß. Eine bessere und genauere Klassifikation wäre dadurch möglich, dass bei der Erstellung einer Seite eine entsprechende Information in den HTML-Code eingefügt würde. Zur Bestimmung des Seitentyps könnte diese Information dann einfach ausgelesen werden.

Ein grundsätzliches Problem bei der Untersuchung von Nutzerverhalten auf Basis von Webserver-Logdaten ist die nicht-vollständige und/oder fehlerbehaftete Rekonstruktion der Nutzersessions. Die Qualität aller darauf basierenden Analysen hängt von der Korrektheit dieser Sessionrekonstruktion ab. Das hier eingesetzte Verfahren ist laut bestehender Untersuchungen relativ zuverlässig und das beste bei Fehlen eindeutiger Identifizierungsmechanismen. Auf Basis der vorliegenden

Logdaten kann allerdings keine Aussage zur Qualität der durchgeföhrten Rekonstruktion gemacht werden. Dieser Unsicherheitsfaktor bleibt.

7.2.2 Nutzermeinungen

Die im Rahmen der empirischen Auswertung der Umfrageergebnisse validierten Hypothesen bestätigen Proposition 2, d. h. Informations- sowie Systemqualität haben einen signifikanten Einfluss auf die hier vorgeschlagenen Kategorien (Wissen, Vertrauen, Risiko und Loyalität) zur Erfassung von individuell wahrgenommenen Auswirkungen der Webportalsnutzung.

Voraussetzung für einen Vergleich der Umfrageergebnisse der Portale ist eine zumindest grobe Übereinstimmung in den Eigenschaften ihrer Nutzer. Die Auswertungsergebnisse des demographischen Teils des Fragebogens deuten diesbezüglich auf eine ausreichende Übereinstimmung hin. Die Daten beinhalten allerdings zu wenige Informationen, um genauere Aussagen ableiten zu können.

Die Umfrageergebnisse zu den verschiedenen Portalen weichen nur in geringem Maße von einander ab. Das kann darauf zurückzuföhren sein, dass bei allen untersuchten Portalen ein ähnliches Qualitätsniveau vorliegt oder aber dass die Anzahl der Stufen (1–5) der auf dem Fragebogen verwendeten Ratingskala für ausreichende Differenzierbarkeit zu gering ist.

Ein häufiges Problem bei schriftlichen Umfragen ist die Nicht-Beantwortung von Fragen. Unvollständige Fragebögen werden hier mittels Multipler Imputation vervollständigt. Untersuchungen haben gezeigt, dass die Multiple Imputation eine valide statistische Auswertung der Daten ermöglicht, als jedes andere Imputationsverfahren oder die Betrachtung nur der vollständigen Fälle. Da jedoch die wahre Verteilung der Daten in der Stichprobe aufgrund der fehlenden Werte nicht bekannt ist, bleibt ein Rest an Unsicherheit.

7.2.3 Aufwendungen

Die Ergebnisse lassen die Annahme zu, dass Portalpflegeaufwand und Portalerfolg in Zusammenhang stehen. Aufgrund der hier erfassten Daten ist somit eine ansatzweise Effizienzbetrachtung in Bezug auf die erzielten Effekte sowie im Vergleich zu den anderen Portalen möglich. Diese ist aber zu relativieren, da das Ergebnis auf Schätzungen der Portalbetreiber und auf Logdaten aus einem recht kurzen Zeitraum basiert.

Mechanismen zur exakten Erfassung der Contentänderungsvorgänge, der wiederkehrenden Besucher sowie die Auswertung von Logdaten aus einem längeren Zeitraum (z. B. ein Jahr) würden zu präziseren Ergebnissen führen und damit genauere Aussagen ermöglichen. Auf entsprechende Möglichkeiten konnte in der vorliegenden Fallstudie nicht zurückgegriffen werden.

Die in Proposition 3 getroffene Annahme, dass Portalpflegevorgänge sich umso bezahlter machen, je mehr Nutzer diese potentiell wahrnehmen, konnte durch die hier empirisch erfassten Zusammenhänge nicht direkt bestätigt werden.

7.2.4 Gesamtbewertung

Die grundlegende Annahme ist, dass das Nutzerverhalten, die Nutzermeinungen sowie die Aufwendungen zur Pflege der Inhalte wichtige und ausreichende Indikatoren sind, um aus deren Betrachtung sinnvolle Aussagen zu Erfolg und Verbesserungspotential eines Portals ableiten zu können. Zwei Argumente für die Gültigkeit dieser Annahme sind zum einen der hohe Stellenwert dieser Kategorien in der Literatur und zum anderen die in der Fallstudie zu diesen Kategorien abgegebenen hohen Präferenzratings seitens des einbezogenen Entscheidungsträgers. Analog dazu kann die Gültigkeit der hier vorgeschlagenen Operationalisierungskriterien für diese drei Kategorien begründet werden.

Da Erfolg aus Sicht unterschiedlicher Stakeholder differenziert bewertet werden kann, ist durch die Berücksichtigung individueller Präferenzen des Entscheidungsträgers im Bewertungsmodell eine Anpassung an Stakeholderbedürfnisse notwendig und möglich.

Der praktische Nutzen des entwickelten Modells konnte in der Fallstudie aufgrund der aus den Ergebnissen gezogenen Schlussfolgerungen durch den als Entscheidungsträger einbezogenen Portalbetreiber gezeigt werden. Die festgelegte Zielsetzung, dass das Modell zur Ermittlung von Erfolgsniveau, Schwachstellen und möglichen Verbesserungsansätzen dienen soll, wurde somit erreicht.

Der Vergleich der Portale (d. h. die Erfolgsbewertung) basiert ausschließlich auf dem verwendeten Satz quantitativer Bewertungskriterien. Erfolgsfaktoren, die nicht quantitativ erfassbar sind, und jene, die nicht Teil des Bewertungsmodells sind, werden nicht berücksichtigt. Dieses kann zu einer unvollständigen Betrachtung eines Portals und damit zu einer den Tatsachen nicht ganz entsprechenden Erfolgsbeurteilung führen.

7.3 Ausblick

Der vorgestellte Ansatz basiert auf der Annahme der Vergleichbarkeit von Nutzerverhalten und Nutzermeinungen zwischen verschiedenen Portalen. Man muss hierbei davon ausgehen, dass die Nutzergruppen der verschiedenen Portale bis zu einem gewissen Grad in ihren Eigenschaften differieren und diese Unterschiede sich auch möglicherweise auf das Verhalten und die Meinungen auswirken. Wie groß der Einfluss unterschiedlicher Nutzereigenschaften diesbezüglich tatsächlich ist, wäre daher eine interessante Fragestellung für zukünftige Untersuchungen. Hierzu könnte man beispielsweise in einem Experiment die Nutzer zweier verschiedener Portale dasselbe benutzen und bewerten lassen und die Unterschiede analysieren.

Die vorliegenden Ergebnisse zu den Nutzermeinungen für die untersuchten Portale differieren nur in geringem Maße. Das kann als ein Hinweis auf ein recht ähnliches Qualitätsniveau bei den verschiedenen Portalen gesehen werden. Die Ursache kann aber auch in einer zu geringen Zahl an Stufen (1–5) bei der verwendeten Ratingskala liegen. In Folgeuntersuchungen könnte ermittelt werden, ob beispielsweise eine 7- oder 10-Stufen-Skala zu differenzierteren Ergebnissen führen würde.

Die Ergebnisse zur Aufwandsbewertung können nur als erster Hinweis gesehen werden. Wie oben beschrieben wäre eine deutlich bessere Bewertung mittels der Auswertung des Verhaltens wiederkehrender Besucher möglich. In Folgeuntersuchungen könnte dafür längerfristig ein Nutzeridentifikationsmechanismus, wie z. B. Cookies, eingesetzt werden.

Die Propositionen 1 und 3 werden in dieser Arbeit nicht empirisch validiert. Entsprechende Analysen könnten zur Überprüfung und ggf. Verbesserung und Erweiterung des entwickelten Modells beitragen.

Zur Validierung von Proposition 1 könnten in einem Experiment Nutzermeinungen und Nutzungsmuster derselben Nutzer erfasst und verglichen werden. So ließe sich ableiten, ob der angenommene Zusammenhang zwischen Nutzungsmustern und individuell wahrgenommenen Auswirkungen bzw. individuell eingeschätztem Portalerfolg besteht.

Zur Validierung von Proposition 3 könnte ein Experiment mit variierender Anzahl an Teilnehmern mehrmals durchgeführt werden. Bei diesem müssten die Teilnehmer auf vorher nicht angekündigte und in unregelmäßigen Abständen durchgeführte Änderungen im Portal mit einer Rückmeldung reagieren. Die Anzahl der Beteiligten an einem Durchlauf und die Menge der Rückmeldungen sollten dementsprechend miteinander korrelieren – vorausgesetzt, die Inhalte des Portals sind ausreichend interessant, um eine regelmäßige Nutzung hervorzurufen. Ein entsprechendes Ergebnis wäre ein Hinweis darauf, dass der Breitungsgrad von Contentänderungen größer ist und diese damit potentiell lohnender sind, je mehr Nutzer eines Portals diese wahrnehmen – Lohnend in Bezug auf das Ziel, dass bereitgestellte Informationen möglichst viele Nutzer erreichen sollen.

Die verwendeten Bewertungskriterien im entwickelten Modell basieren auf in der Literatur verbreiteten Ansätzen. Dazu wurde eine Auswahl an für den hier betrachteten Rahmen als relevant erachteten Kriterien getroffen, ohne jedoch Anspruch auf Vollständigkeit zu erheben. Zusammenfassend lässt sich sagen, dass das Modell einen ersten Schritt in Richtung eines standardisierten Vorgehens zur Bewertung der Qualität von Informationen und der Informationsbereitstellung auf nicht-kommerziellen Portalen darstellt und dementsprechend als Basis für zukünftige Forschung in dieser Richtung dienen kann.

Literaturverzeichnis

Agarwal u. Venkatesh 2002

AGARWAL, Ritu ; VENKATESH, Viswanath: Assessing a Firm's Web Presence: A Heuristic Evaluation Procedure for the Measurement of Usability. In: *Information Systems Research* 13 (2002), Nr. 2, S. 168–186

Agrawal u. Srikant 1995

AGRASWAL, Rakesh ; SRIKANT, Ramakrishnan: Mining sequential patterns. In: *Proceedings of International Conference on Data Engineering*. Taipeh, Taiwan, 1995

Aitkenhead 2005

AITKENHEAD, Tony: Web Portals in Government Service. In: TATNALL, Arthur (Hrsg.): *Web Portals: The New Gateways to Internet Services*. London : Idea Group Publishing, 2005, S. 212–229

Aladwani u. Palvia 2002

ALADWANI, Adel M. ; PALVIA, Prashant C.: Developing and validating an instrument for measuring user-perceived web quality. In: *Information & Management* 39 (2002), S. 467–476

Albrecht 2001

ALBRECHT, Karl: Information Quality. In: *Executive Excellence* (2001), August

Alexander u. Tate 1999

ALEXANDER, Janet E. ; TATE, Marsha A.: *Web Wisdom: How to Evaluate and Create Information Quality on the Web*. London : Lawrence Erlbaum Associates, 1999

Appan u. Mellarkod 2003

APPAN, Radha ; MELLARKOD, Vidhya: The Impact of Website Design Effectiveness on Customer's Trusting Beliefs About a Pure Internet Retailer: A Conceptual Study. In: *Proceedings of the Ninth Americas Conference on Information Systems*, 2003

Bailey u. Pearson 1983

BAILEY, J. E. ; PEARSON, S. W.: Development of a Tool for Measuring and Analyzing Computer User Satisfaction. In: *Management Science* 29 (1983), Nr. 5, S. 530–545

Bajec 2005

BAJEC, Marko: Educational Portals: A Way to Get an Integrated, User-Centric University Information System. In: TATNALL, Arthur (Hrsg.): *Web Portals: The New Gateways to Internet Services*. London : Idea Group Publishing, 2005, S. 252–269

Balasubramanian u. a. 2003

BALASUBRAMANIAN, Sridhar ; KONANA, Prabhudev ; MENON, Nirup M.: Customer Satisfaction in Virtual Environments: A Study of Online Investing. In: *Management Science* 49 (2003), Nr. 7, S. 871–889

Bamberg u. Coenenberg 2002

BAMBERG, Günter ; COENENBERG, Adolf G.: *Betriebswirtschaftliche Entscheidungslehre*. 11. München : Vahlen, 2002

Bart u. a. 2005

BART, Yakov ; SHANKAR, Venkatesh ; SULTAN, Fareena ; URBAN, Glen L.: Are the Drivers and Role of Online Trust the Same for all Web Sites and Consumers? A Large-Scale Exploratory Empirical Study. In: *Journal of Marketing* 69 (2005), S. 133–152

Benbasat u. Zmud 2003

BENBASAT, Izak ; ZMUD, Robert W.: The identity crisis within the IS discipline: defining and communicating the discipline's core properties. In: *MIS Quarterly* 27 (2003), Nr. 2, S. 183–194

Benbunan-Fich 2001

BENBUNAN-FICH, Raquel: Using protocol analysis to evaluate the usability of a commercial web site. In: *Information & Management* 39 (2001), S. 151–163

Berendt u. a. 2001

BERENDT, B. ; MOBASHER, B. ; SPILOPOULOU, M. ; WILTSIRE, J.: Measuring the Accuracy of Sessionizers for Web Usage Analysis. In: *Workshop on Web Mining at the First SIAM International Conference on Data Mining*. Chicago, IL, USA, April 2001, S. 7–14

Berendt 2002

BERENDT, Bettina: Detail and Context in Web Usage Mining: Coarsening and Visualizing Sequences. In: *WEBKDD '01: Revised Papers from the Third International Workshop on Mining Web Log Data Across All Customers Touch Points*. London, UK : Springer-Verlag, 2002. – ISBN 3-540-43969-2, S. 1–14

Berendt u. Spiliopoulou 2000

BERENDT, Bettina ; SPILOPOULOU, Myra: Analysis of navigation behaviour in web sites integrating multiple information systems. In: *The VLDB Journal* 9 (2000), S. 56–75

Berthon u. a. 1996

BERTHON, Pierre ; PITT, Leyland F. ; WATSON, Richard T.: The World Wide Web as an Advertising Medium: Toward an Understanding of Conversion Efficiency. In: *Journal of Advertising Research* (1996), January/February

Bhimani 1996

BHIMANI, Anish: Securing the commercial Internet. In: *Communications of the ACM* 39 (1996), Nr. 6, S. 29–35

Bortz u. Döring 2006

BORTZ, Jürgen ; DÖRING, Nicola: *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. 4. Heidelberg : Springer, 2006

Brynjolfson 1993

BRYNJOLFSON, E.: The Productivity Paradox of Information Technology. In: *Communications of the ACM* 35 (1993), Nr. 12, S. 66–77

Büchner u. Mulvenna 1998

BÜCHNER, Alex G. ; MULVENNA, Maurice D.: Discovering Internet Marketing Intelligence through Online Analytical Web Usage Mining. In: *SIGMOD Record* 27 (1998), Nr. 4, S. 54–61

Butters 2003

BUTTERS, Geoff: ‘What features in a Portal?’. In: *ARIADNE* (2003), Nr. 35

Cameron 1980

CAMERON, Kim S.: Critical Questions in Assessing Organizational Effectiveness. In: *Organizational Dynamics* 9 (1980), Nr. 2, S. 66–80

Camp 1994

CAMP, Robert C.: *Benchmarking*. München : Hanser, 1994

Cao u. a. 2004

CAO, Mei ; ZHANG, Qingyu ; SEYDEL, John: Measuring E-Commerce Web Site Quality: An Empirical Examination. In: *Proceedings of the Tenth Americas Conference on Information Systems*. New York, NY, USA, 2004

Catledge u. Pitkow 1995

CATLEDGE, Lara D. ; PITKOW, James E.: Characterizing Browsing Strategies in the World-Wide-Web. In: *Computer Networks and ISDN Systems* 27 (1995), Nr. 6

Chan 2000

CHAN, Yolande E.: IT Value: The Great Divide Between Qualitative and Quantitative and Individual and Organizational Measures. In: *Journal of Management Information Systems* 16 (2000), Nr. 4, S. 225–261

Chandrinos u. Trahanias 1998

CHANDRINOS, K. V. ; TRAHANIAS, P. E.: Beyond HTML: Web-Based Information Systems. In: *Sixth DELOS Workshop - Preservation of Digital Information*. Tomar, Portugal, 1998

Chin u. a. 1988

CHIN, John P. ; DIEHL, Virginia A. ; NORMAN, Kent L.: Development of a Tool Measuring User Satisfaction of the Human-Computer Interface. In: *CHI '88 Conference Proceedings: Human Factors in Computing Systems*. New York : ACM, 1988, S. 213–218

Chin 1998

CHIN, Wynne E.: The Partial Least Squares Approach for Structural Equation Modeling. In: MARCOULIDES, George A. (Hrsg.): *Modern Methods for Business Research*. London : Lawrence Erlbaum Associates, 1998

Chung u. Tan 2004

CHUNG, Janine ; TAN, Felix B.: Antecedents of perceived playfulness: an exploratory study on user acceptance of general information-searching websites. In: *Information & Management* 41 (2004), S. 869–881

Cohen u. a. 1998

COHEN, Edith ; KRISHNAMURTHY, Balachander ; REXFORD, Jennifer: Improving End-to-End Performance of the web Using Server Volumes and Proxy Filters. In: *Proceedings of the ACM SIGCOMM*. Vancouver, Kanada, 1998, S. 241–253

Cooley u. a. 1999

COOLEY, R. W. ; MOBASHER, B. ; SRIVASTAVA, J.: Data Preparation for Mining World Wide Web Browsing Patterns. In: *Jurnal of Knowledge and Information Systems* 1 (1999), Nr. 1, S. 5–32

Cooley 2003

COOLEY, Robert: The Use of Web Structure and Content to Identify Subjective Interesting Web Usage Patterns. In: *ACM Transactions on Internet Technology* 3 (2003), Nr. 2, S. 93–116

Corritore u. a. 2005

CORRITORE, Cynthia L. ; MARBLE, Robert P. ; WIEDENBECK, Susan ; KRACHER, Beverly ; CHANDRAN, Ashwin: Measuring Online Trust of Websites: Credibility, Perceived Ease of Use, and Risk. In: *Proceedings of the Eleventh Americas Conference on Information Systems*. Omaha, NE, USA, 2005

Cronbach 1951

CRONBACH, Lee J.: Coefficient alpha and the internal structure of tests. In: *Psychometrika* 16 (1951), Nr. 3, S. 297–334

Cronin u. Taylor 1992

CRONIN, J. J. ; TAYLOR, Steven A.: Measuring Service Quality: A Reexamination and Extension. In: *Journal of Marketing* 56 (1992), Nr. 3, S. 55–68

Cushing 1990

CUSHING, Barry E.: Frameworks, Paradigms, and Scientific Research in Management Information Systems. In: *Journal of Information Systems* 4 (1990), Nr. 2

Cutler u. Sterne 2000

CUTLER, Matt ; STERNE, Jim: *E-Metrics - Business Metrics For The New Economy*. Version: 2000. www.netgen.com. – Online-Ressource. – Letzter Zugriff: 08.05.2004

D'Ambra u. Rice 2001

D'AMBRA, John ; RICE, Ronald E.: Emerging factors in user evaluation of the World Wide Web. In: *Information & Management* 38 (2001), S. 373–383

D'Ambra u. Wilson 2004

D'AMBRA, John ; WILSON, Concepción S.: Explaining perceived performance of the World Wide Web: uncertainty and the task-technology fit model. In: *Internet Research* 14 (2004), Nr. 4, S. 294–310

David u. a. 2002

DAVID, Julie S. ; SCHUFF, David ; LOUIS, Robert S.: Managing Your Total Cost of Ownership. In: *Communications of the ACM* 45 (2002), Nr. 1, S. 101–106

Davis 1986

DAVIS, Fred D.: *A Technology Acceptance Model for Empirically Testing New End-user Information Systems Theory and Results*, Massachusetts Institute of Technology, Diss., 1986

Davis 1989

DAVIS, Fred D.: Perceived usefulness, perceived ease-of-use, and user acceptance of information technology. In: *MIS Quarterly* (1989), September, S. 319–340

Davis 1997

DAVIS, Gordon B. (Hrsg.): *The Blackwell Encyclopedic Dictionary of Management Information Systems*. Oxford : Blackwell, 1997

Davis u. Olson 1985

DAVIS, Gordon B. ; OLSON, Margrethe H.: *Management Information Systems: Conceptual Foundations, Structure and Development*. 2. Edition. New York : McGraw Hill, 1985

Davison u. Hinkley 1999

DAVISON, A. C. ; HINKLEY, D. V.: *Bootstrap Methods and their Application*. Cambridge University Press, 1999

Deane u. a. 1998

DEANE, Frank P. ; PODD, John ; HENDERSON, Ron D.: Relationship Between Self-report and Log Data Estimates of Information System Usage. In: *Computers in Human Behavior* 14 (1998), Nr. 4, S. 621–636

Delić 2007

DELIĆ, Daniel: A Multiple-Criteria Decision Model for Web Portal Success Evaluation. In: KÖPPEN, Veit (Hrsg.) ; MÜLLER, Roland M. (Hrsg.): *Business Intelligence: Methods and Applications*. Hamburg : Dr. Kovač, 2007, S. 175–185

Delić u. Lenz 2006

DELIĆ, Daniel ; LENZ, Hans-J.: Towards Success Evaluation: Multiple-Criteria Analysis of Benefits and Costs for Non-Profit Web Portals. In: *Proceedings of the Twelfth Americas Conference on Information Systems*. Acapulco, Mexico, 2006, S. 512–516

Delić u. Lenz 2008a

DELIĆ, Daniel ; LENZ, Hans-J.: Benchmarking User Perceived Impact for Web Portal Success Evaluation. In: *Journal of Information and Organizational Sciences* 32 (2008), Nr. 1, S. 1–14

Delić u. Lenz 2008b

DELIĆ, Daniel ; LENZ, Hans-J.: Non-Profit Web Portals - Usage based Benchmarking for Success Evaluation. In: PREISACH, Christine (Hrsg.) ; BURKHARDT, Hans (Hrsg.) ; SCHMIDT-THIEME, Lars (Hrsg.) ; DECKER, Reinhold (Hrsg.): *Data Analysis, Machine Learning and Applications. Proceedings of the 31st Annual Conference of the Gesellschaft für Klassifikation e. V. 2007*. Berlin : Springer, 2008, S. 561–568

DeLone u. McLean 1992

DELONE, William H. ; MCLEAN, Ephraim R.: Information Systems Success: The Quest for the Dependent Variable. In: *Information Systems Research* 3 (1992), Nr. 1, S. 60–95

DeLone u. McLean 2003

DELONE, William H. ; MCLEAN, Ephraim R.: The DeLone and McLean Model of Information Systems Success: A Ten-Year Update. In: *Journal of Management Information Systems* 19 (2003), Nr. 4, S. 9–30

Devaraj u. a. 2002

DEVARAJ, Sarv ; FAN, Ming ; KOHLI, Rajiv: Antecedents of B2C Channel Satisfaction and Preference: Validating e-Commerce Metrics. In: *Information Systems Research* 13 (2002), Nr. 3, S. 316–333

Dhyani u. a. 2002

DHYANI, Devanshu ; NG, Wee K. ; BHOWMICK, Sourav S.: A survey of Web metrics. In: *ACM Computing Surveys* 34 (2002), Nr. 4, S. 469–503

Dishaw u. Strong 1999

DISHAW, Mark T. ; STRONG, Diane M.: Extending the technology acceptance model with task-technology fit constructs. In: *Information & Management* 36 (1999), Nr. 1

Doll u. Torkzadeh 1988

DOLL, W. J. ; TORKZADEH, G.: The Measurement of End-User Computing Satisfaction. In: *MIS Quarterly* 12 (1988), Nr. 2, S. 259–274

Drèze u. Zufryden 1997

DRÈZE, Xavier ; ZUFRYDEN, Fred: Testing Web Site Design and Promotional Content. In: *Journal of Advertising Research* (1997), S. 77–91

Dunn u. Varano 1999

DUNN, Jaquelin R. ; VARANO, Michael W.: Leveraging web-based information systems. In: *Information Systems Management* 16 (1999), Nr. 4, S. 60–69

Elmuti u. a. 1997

ELMUTI, Dean ; KATHAWALA, Yunus ; LLOYED, Scott: The Benchmarking Process: Assessing Its Value and Limitations. In: *Industrial Management* (1997), S. 12–19

Ester u. Sander 2000

ESTER, Martin ; SANDER, Jörg: *Knowledge Discovery in Databases: Techniken und Anwendungen*. Heidelberg : Springer Verlag, 2000

Etezadi-Amoli u. Farhoomand 1996

ETEZADI-AMOLI, Jamshid ; FARHOOMAND, Ali F.: A structural model of end user computing satisfaction and user performance. In: *Information & Management* 30 (1996), S. 65–73

Facca u. Lanzi 2005

FACCA, Federico M. ; LANZI, Pier L.: Mining interesting knowledge from weblogs: a survey. In: *Data & Knowledge Engineering* 53 (2005), S. 225–241

Figueira u. a. 2005

FIGUEIRA, José (Hrsg.) ; GRECO, Salvatore (Hrsg.) ; EHRGOTT, Matthias (Hrsg.): *Multiple Criteria Decision Analysis: State of the Art Surveys*. Boston : Springer Science + Business Media, 2005

Fishbein u. Ajzen 1975

FISHBEIN, Martin ; AJZEN, Icek: *Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research*. MA : Addison-Wesley, 1975

Flavián u. a. 2006

FLAVIÁN, Carlos ; GUINALÍU, Miguel ; GURREA, Raquel: The role played by perceived usability, satisfaction and consumer trust on website loyalty. In: *Information & Management* 43 (2006), S. 1–14

Flesca u. a. 2005

FLESCA, Sergio ; GRECO, Sergio ; TAGARELLI, Andrea ; ZUMPANO, Ester: Mining User Preferences, Page Content and Usage to Personalize Website Navigation. In: *World Wide Web: Internet and Web Information Systems* 8 (2005), Nr. 3, S. 317–345

Fogg u. a. 2002

FOGG, B. J. ; SOOHOO, Cathy ; DANIELSON, David ; MARABLE, Leslie ; STANFORD, Julianne ; TAUBER, Ellen R.: How Do People Evaluate a Web Site's Credibility? / Persuasive Technology Lab, Stanford University. 2002. – Forschungsbericht

Fornell 1992

FORNELL, Claes: A National Customer Satisfaction Barometer: The Swedish Experience. In: *Journal of Marketing* 56 (1992), S. 6–21

Fornell u. Bookstein 1982

FORNELL, Claes ; BOOKSTEIN, Fred L.: Two Structural Equation Models: LISREL and PLS Applied to Consumer Exit-Voice Theory. In: *Journal of Marketing Research* 19 (1982), S. 440–452

Fornell u. a. 1996

FORNELL, Claes ; JOHNSON, Michael D. ; ANDERSON, Eugene W. ; CHA, Jaesung ; BRYANT, Barbara E.: The American Customer Satisfaction Index: Nature, Purpose, and Findings. In: *Journal of Marketing* 60 (1996), S. 7–18

Frawley u. a. 1992

FRAWLEY, W. J. ; PIATETSKY-SHAPIRO, G. ; MATHEUS, C. J.: Knowledge discovery in databases - an overview. In: *Ai Magazine* 13 (1992), S. 57–70

Freeman 1984

FREEMAN, R. E.: *Strategic management: A stakeholder approach*. Boston : Pitman, 1984

French 1988

FRENCH, S.: *Decision Theory – An Introduction to the Mathematics of Rationality*. New York : Springer, 1988

Fu u. a. 1999

FU, Y. ; SANDHU, K. ; SHIH, M.: Clustering of Web Users Based on Access Patterns. In: *Proceedings of the 1999 KDD Workshop on Web Mining*. San Diego, CA, USA, 1999

Fu u. a. 2001

FU, Yongjian ; CREADO, Mario ; JU, Chunhua: Reorganizing web sites based on user access patterns. In: *CIKM '01: Proceedings of the Tenth International Conference on Information and Knowledge Management*. New York, NY, USA : ACM Press, 2001, S. 583–585

Fu u. a. 2000

Kapitel A Generalization-Based Approach to Clustering of Web Usage Sessions.
In: FU, Yongjian ; SANDHU, Kanwalpreet ; SHIH, Ming-Yi: *Web Usage Analysis and User Profiling*. Berlin : Springer, 2000 (Lecture Notes in Computer Science)

Gallaugh u. Downing 2005

GALLAUGHER, John M. ; DOWNING, Charles E.: Portal Combat Revisited: Success Factors and Evolution in Consumer Web Portals. In: TATNALL, Arthur (Hrsg.): *Web Portals: The New Gateways to Internet Services*. London : Idea Group Publishing, 2005, S. 40–63

Galletta u. a. 2004

GALLETTA, Dennis F. ; HENRY, Raymond ; MCCOY, Scott ; POLAK, Peter: Web Site Delays: How Tolerant are Users? In: *Journal of the AIS* 5 (2004), Nr. 1, S. 1–28

Galliers 1992

GALLIERS, Robert D.: Choosing Information Systems Research Approaches. In: GALLIERS, Robert (Hrsg.): *Information Systems Research: Issues, Methods and Practical Guidelines*. Oxford : Blackwell, 1992, Kapitel 8

Gartner 2007

Gartner Says EMEA Portals, Process and Middleware Software Market Revenue Increased 16 percent in 2006. www.gartner.com/it/page.jsp?id=507724.
Version: 2007. – Letzter Zugriff: 05.05.2008

Gefen 2002a

GEFEN, David: Customer Loyalty in E-Commerce. In: *Journal of the AIS* 3 (2002), S. 27–51

Gefen 2002b

GEFEN, David: Reflections on the Dimensions of Trust and Trustworthiness among Online Consumers. In: *The DATA BASE for Advances in Information Systems* 33 (2002), Nr. 3, S. 38–53

Gefen u. a. 2003

GEFEN, David ; KARAHANNA, Elena ; STRAUB, Detmar W.: Trust and TAM in Online Shopping: An Integrated Model. In: *MIS Quarterly* 27 (2003), Nr. 1, S. 51–90

Gefen u. Straub 2000

GEFEN, David ; STRAUB, Detmar: The Relative Importance of Perceived Ease of Use in IS Adoption: A Study of E-Commerce Adoption. In: *Journal of the AIS* 1 (2000), Nr. 8

Gefen u. Straub 2005

GEFEN, David ; STRAUB, Detmar: A Practical Guide to Factorial Validity using PLS-Graph: Tutorial and annotated Example. In: *Communications of the AIS* 16 (2005), S. 91–109

Gefen u. a. 2000

GEFEN, David ; STRAUB, Detmar W. ; BOUDREAU, Marie-Claude: Structural Equation Modeling and Regression: Guidelines for Research Practice. In: *Communications of the AIS* 4 (2000), Nr. 7

Gehrke u. Turban 1999

GEHRKE, Dave ; TURBAN, Efraim: Determinants of Successful Website Design: Relative Importance and Recommendations for Effectiveness. In: *Proceedings of the 32nd Hawaii International Conference on System Sciences*, 1999

Gillenson u. a. 2000

GILLENSON, Mark L. ; SHERRELL, Daniel L. ; CHEN, Lei-da: A Taxonomy of Web Site Traversal Patterns and Structures. In: *Communications of the AIS* 3 (2000), Nr. 17

Goodhue u. Thompson 1995

GOODHUE, D. ; THOMPSON, R.: Task-technology fit and individual performance. In: *MIS Quarterly* 19 (1995), Nr. 2, S. 213–236

Goodhue 1988

GOODHUE, Dale: I/S attitudes: toward theoretical and definitional clarity. In: *SIGMIS Database* 19 (1988), Nr. 3-4, S. 6–15

Goodhue u. a. 2000

GOODHUE, Dale L. ; KLEIN, Barbara D. ; MARCH, Salvatore T.: User evaluations of IS as surrogates for objective performance. In: *Information & Management* 38 (2000), S. 87–101

Grandon u. Ranganathan 2001

GRANDON, Elizabeth E. ; RANGANATHAN, C.: The Impact of Content and Design of Web Sites on Online Sales. In: *Proceedings of the Seventh Americas Conference on Information Systems*. Boston, MA, USA, 2001

Grigoroudis u. a. 2008

GRIGOROUDIS, Evangelos ; LITOS, Charalambos ; MOUSTAKIS, Vassilis A. ; POLITIS, Yannis ; TSIRONIS, Loukas: The assessment of user-perceived web quality: Application of a satisfaction benchmarking approach. In: *European Journal of Operational Research* 187 (2008), Nr. 3, S. 1346–1357

Grothe u. Gentsch 2000

GROTHE, Martin ; GENTSCH, Peter: *Business Intelligence*. München : Addison-Wesley, 2000

Grover u. a. 1996

GROVER, Varun ; JEONG, Seung R. ; SEGARS, Albert H.: Information Systems Effectiveness: The Construct Space and Patterns of Application. In: *Information & Management* 31 (1996), S. 177–191

Gupta u. Kim 2004

GUPTA, Sumeet ; KIM, Hee-Woong: Virtual Community: Concepts, Implications, and Future Research Directions. In: *Proceedings of the Tenth Americas Conference on Information Systems*. New York, NY, USA, 2004, S. 2679–2687

Hallam-Baker u. Behlendorf 2002

HALLAM-BAKER, Phillip M. ; BEHLENDORF, Brian: *W3C Working Draft WD-logfile-960323: Extended Log File Format.* www.w3.org/TR/WD-logfile.html. Version: 2002. – Letzter Zugriff: 18.11.2006

Hamilton u. Ives 1992

HAMILTON, Scott ; IVES, Blake: MIS Research Strategies. In: GALLIERS, Robert (Hrsg.): *Information Systems Research: Issues, Methods and Practical Guidelines*. Oxford : Blackwell, 1992, Kapitel 7

van der Heijden 2003

HEIJDEN, Hans van d.: Factors influencing the usage of websites: the case of a generic portal in The Netherlands. In: *Information & Management* 40 (2003), S. 541–549

Heitjan u. Rubin 1990

HEITJAN, D. F. ; RUBIN, D. B.: Inference From Coarse Data via Multiple Imputation With Application to Age Heaping. In: *Journal of the American Statistical Association* 85 (1990), S. 304–314

Hightower u. a. 1998

HIGHTOWER, Christy ; SIH, Julie ; TILGHMAN, Adam: Recommendations for Benchmarking Web Site Usage among Academic Libraries. In: *College & Research Libraries* 59 (1998), Nr. 1, S. 61–79

Hoffman u. a. 1999

HOFFMAN, Donna L. ; NOVAK, Thomas P. ; PERALTA, Marcos: Building Consumer Trust Online. In: *Communications of the ACM* 42 (1999), Nr. 4

Hong u. a. 2004

HONG, Weiyin ; THONG, James Y. L. ; TAM, Kar Y.: Does Animation Attract Online Users' Attention? The Effects of Flash on Information Search Performance and Perceptions. In: *Information Systems Research* 15 (2004), Nr. 1, S. 60–87

Hornbæk 2006

HORNBÆK, Kasper: Current practice in measuring usability: Challenges to usability studies and research. In: *International Journal of Human-Computer Studies* 64 (2006), Nr. 2, S. 79–102

Huizingh 2000

HUIZINGH, Eelko K.: The content and design of web sites: an empirical study. In: *Information & Management* 37 (2000), S. 123–134

Hwang u. Yi 2002

HWANG, Yujong ; YI, Mun Y.: Predicting the Use of Web-Based Information Systems: Intrinsic Motivation and Self-Efficacy. In: *Proceedings of the Eighth Americas Conference on Information Systems*, 2002, S. 1076–1081

Igbaria u. Tan 1997

IGBARIA, M. ; TAN, M.: The consequences of information technology acceptance on subsequent individual performance. In: *Information & Management* 32 (1997), S. 113–121

Irani 2002

IRANI, Zahir: Information systems evaluation: navigating through the problem domain. In: *Information & Management* 40 (2002), S. 11–22

ISO 9241-11 1998

Ergonomic requirements for office work with visual display terminals (VDTs) – Part 11: Guidance on usability. www.iso.org. Version: 1998

Ives u. a. 1983

IVES, Blake ; OLSON, Margrethe H. ; BAROUDI, Jack J.: The Measurement of User Information Satisfaction. In: *Communications of the ACM* 26 (1983), Nr. 10, S. 785–793

Ivory u. Hearst 2001

IVORY, Melody Y. ; HEARST, Marti A.: The State of the Art in Automating Usability Evaluation of User Interfaces. In: *ACM Computing Surveys* 33 (2001), Nr. 4, S. 470–516

Ivory u. a. 2001

IVORY, Melody Y. ; SINHA, Rashmi R. ; HEARST, Marti A.: Empirically Validated Web Page Design Metrics. In: *CHI '01: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York : ACM, 2001, S. 53–60

Jacobson 2000

JACOBSON, Carl: New Horizons: Institutional Portals. In: *Educause Review* 35 (2000), Nr. 4, S. 58–59

Jutla u. a. 1999

JUTLA, Dawn ; BODORIK, Peter ; WANG, Yie: Developing internet e-commerce benchmarks. In: *Information Systems* 24 (1999), Nr. 6, S. 475–493

Kaylor u. a. 2001

KAYLOR, Charles ; DESHAZO, Randy ; ECK, David V.: Gauging e-government: A report on implementing services among American cities. In: *Government Information Quarterly* 18 (2001), S. 293–307

Keeker 1997

KEEKER, Kevin: *Improving Web Site Usability and Appeal: Guidelines compiled by MSN Usability Research.* msdn.microsoft.com/library/default.asp?url=/library/en-us/dnsiteplan/html/improvingsiteusa.asp. Version: 1997. – Letzter Zugriff: 21.10.2004

Keen 1980

KEEN, Peter G.: MIS Research: Reference Disciplines and a Cumulative Tradition. In: *Proceedings of the First International Conference on Information Systems*, 1980, S. 9–18

Kettinger u. Lee 1997

KETTINGER, William J. ; LEE, Choong C.: Pragmatic Perspectives on the Measurement of Information Systems Service Quality. In: *MIS Quarterly* (1997), Juni, S. 223–239

Kim 2000

KIM, Euijin: EBQUAL: A Success Model of E-business. In: *Proceedings of the Sixth Americas Conference on Information Systems*. Long Beach, CA, USA, 2000

King u. a. 2001

KING, Gary ; HONAKER, James ; JOSEPH, Anne ; SCHEVE, Kenneth: Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation. In: *American Political Science Review* 95 (2001), Nr. 1, S. 49–69

Kirchhof u. a. 2004

KIRCHHOF, Anja ; GURZKI, Thorsten ; HINDERER, Henning ; VLACHAKIS, Joannis: *Was ist ein Portal? Definition und Einsatz von Unternehmensportalen.* www.gurzki.de/publications/padem/Whitepaper%20Was%20ist%20ein%20Portal%20Gurzki.pdf. Version: 2004. – Whitepaper Fraunhofer IAO, Letzter Zugriff: 10.11.2004

Klopping u. McKinney 2004

KLOPPING, Inge M. ; MCKINNEY, Earl: Extending the Technology Acceptance Model and the Task-Technology Fit Model to Consumer E-Commerce. In: *Information Technology, Learning, and Performance Journal* 22 (2004), Nr. 1, S. 35–48

Knight u. Burn 2005

KNIGHT, Shirlee-Ann ; BURN, Janice: Developing a Framework for Assessing Information Quality on the World Wide Web. In: *Informing Science Journal* 8 (2005), S. 159–172

König 1994

KÖNIG, Wolfgang: Mitteilungen der Wissenschaftlichen Kommission Wirtschaftsinformatik. Profil der Wirtschaftsinformatik. In: *Wirtschaftsinformatik* 36 (1994), Nr. 1, S. 80–81

Kosala u. Blockeel 2000

KOSALA, Raymond ; BLOCKEEL, Hendrik: Web Mining Research: A Survey. In: *SIGKDD Explorations: Newsletter of the Special Interest Group (SIG) on Knowledge Discovery & Data Mining, ACM* 2 (2000)

Koufaris 2002

KOUFARIS, Marios: Applying the Technology Acceptance Model and Flow Theory to Online Consumer Behavior. In: *Information Systems Research* 13 (2002), Nr. 2, S. 205–223

Kralisch 2005

KRALISCH, Anett: *The Impact of Culture and Language on the Use of the Internet: Empirical Analyses of Behaviour and Attitudes*, Humboldt Universität zu Berlin, Diss., 2005

Kwai Fun IP u. Wagner 2008

KWAI FUN IP, Rachael ; WAGNER, Christian: Weblogging: A study of social computing and its impact on organizations. In: *Decision Support Systems* 45 (2008), Nr. 2, S. 242–250

Lala u. a. 2002

LALA, Vishal ; ARNOLD, Vicky ; SUTTON, Steve G. ; GUAN, Liming: The impact of relative information quality of e-commerce assurance seals on Internet purchasing behavior. In: *International Journal of Accounting Information Systems* 3 (2002), S. 237–253

Landrum u. Prybutok 2004

LANDRUM, Hollis ; PRYBUTOK, Victor R.: A service quality and success model for the information service industry. In: *European Journal of Operations Research* 156 (2004), S. 628–642

Larsen 2003

LARSEN, Kai R. T.: A Taxonomy of Antecedents of Information Systems Success: Variable Analysis Studies. In: *Journal of Management Information Systems* 20 (2003), Nr. 2, S. 169–246

Lavoie u. Nielsen 1999

LAVOIE, Brian ; NIELSEN, Henrik F.: *Web Characterization Terminology & Definitions Sheet.* www.w3.org/1999/05/WCA-terms/. Version: 1999. – Letzter Zugriff: 20.11.2006

Lee u. a. 2007

LEE, Ching-Chang ; CHENG, Hsing K. ; CHENG, Hui-Hsin: An empirical study of mobile commerce in insurance industry: Task-technology fit and individual differences. In: *Decision Support Systems* 43 (2007), S. 95–110

Lee u. a. 2001

LEE, Juhnyoung ; PODLASECK, Mark ; SCHONBERG, Edith ; HOCH, Robert: Visualization and Analysis of Clickstream Data of Online Stores for Understanding Web Merchandising. In: *Data Min. Knowl. Discov.* 5 (2001), Nr. 1-2, S. 59–84

Lee u. a. 2000a

LEE, Juhnyoung ; PODLASECK, Mark ; SCHONBERG, Edith ; HOCH, Robert ; GOMORY, Stephen: Analysis and Visualization of Metrics for Online Merchandising. In: *WEBKDD '99: Revised Papers from the International Workshop on Web Usage Analysis and User Profiling*. London, UK : Springer-Verlag, 2000, S. 126–141

Lee u. a. 2000b

LEE, Jungwon ; KIM, Jinwoo ; MOON, Jae Y.: What makes Internet users visit cyber stores again? Key design factors for customer loyalty. In: *CHI '00: Proceedings of the SIGCHI conference on Human factors in computing systems*. New York : ACM, 2000, S. 305–312

Lee u. a. 2002

LEE, Yang W. ; STRONG, Diane M. ; KAHN, Beverly K. ; WANG, Richard Y.: AIMQ: a methodology for information quality assessment. In: *Information & Management* 40 (2002), S. 133–146

Lee u. a. 2003

LEE, Younghwa ; KOZAR, Kenneth A. ; LARSEN, Kai R.: The Technology Acceptance Model: Past, Present, and Future. In: *Communications of the AIS* 50 (2003), S. 752–780

Lewis 1993

LEWIS, James R.: IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use / IBM. Boca Raton, FL, USA, 1993. – Forschungsbericht

Li 2004

LI, Honglei: Virtual Community Studies: A Literature Review, Synthesis and Research Agenda. In: *Proceedings of the Tenth Americas Conference on Information Systems*. New York, NY, USA, 2004

Likert 1932

LIKERT, R.: A Technique for the Measurement of Attitudes. In: *Archives of Psychology* 140 (1932), S. 1–55

Lin u. Wu 2002

LIN, Cathy S. ; WU, Sheng: Exploring the Impact of Online Service Quality on Portal Site Usage. In: *Proceedings of the 35th Hawaii International Conference on System Sciences*, 2002

Little u. Rubin 2004

LITTLE, Roderick J. A. ; RUBIN, Donald B.: *Statistical Analysis with Missing Data*. 2. Hoboken, NJ, USA : Wiley-Interscience, 2004

Liu u. Arnett 2000

LIU, Chang ; ARNETT, Kirk P.: Exploring the factors associated with Web site success in the context of electronic commerce. In: *Information & Management* 38 (2000), S. 22–33

Looney u. Lyman 2000

LOONEY, Michael ; LYMAN, Peter: Portals in Higher Education: What Are They, and What is Their Potential. In: *Educause Review* 35 (2000), Nr. 4, S. 28–37

Lowry u. a. 2000

LOWRY, Paul B. ; ANDERSON, H. W. ; WILSON, Dennis C. ; LEROY, Gondy ; LIN, Lin: MIS Legitimacy and the Proposition of a New Multidimensional Model of MIS. In: *3rd Annual Conference of the Southern Association for Information Systems (SAIS)*. Atlanta, GA, USA, 2000

van Maanen u. Berghout 2002

MAANEN, Henno van ; BERGHOUT, Egon: Cost management of IT beyond cost of ownership models: a state of the art overview of the Dutch financial services industry. In: *Evaluation and Program Planning* 25 (2002), S. 167–173

Mason 1978

MASON, Richard O.: Measuring Information Output: A Communication Systems Approach. In: *Information & Management* 5 (1978), Nr. 1, S. 219–234

Mayer u. a. 1995

MAYER, Roger C. ; DAVIS, James H. ; SCHOORMAN, F. D.: An Integrative Model of Organizational Trust. In: *Academy of Management Review* 20 (1995), Nr. 3, S. 709–734

McKinney u. a. 2002

MCKINNEY, Vicki ; YOON, Kanghyun ; ZAHEDI, Fatemeh M.: The Measurement of Web-Customer Satisfaction: An Expectation and Disconfirmation Approach. In: *Information Systems Research* 13 (2002), Nr. 3, S. 296–315

McKnight u. a. 2002

MCKNIGHT, D. H. ; CHOUDHURY, Vivek ; KACMAR, Charles: Developing and Validating Trust Measures for e-Commerce: An Integrative Typology. In: *Information Systems Research* 13 (2002), Nr. 3, S. 334–359

Mill 1865

MILL, John S.: *August Comte and Positivism*. – Letzter Zugriff: 28.12.2007.
www.gutenberg.org/etext/16833

Miller 2003

MILLER, Paul: Towards a Typology for Portals. In: *ARIADNE* (2003), Nr. 37.
<http://www.ariadne.ac.uk/issue37/miller/>. – Letzter Zugriff: 11.11.2004

Mingers 2001

MINGERS, John: Combining IS Research Methods: Towards a Pluralist Methodology. In: *Information Systems Research* 12 (2001), Nr. 3, S. 240–259

Mirani u. Lederer 1998

MIRANI, Rajesh ; LEDERER, Albert L.: An instrument for assessing the organizational benefits of IS projects. In: *Decision Sciences* 29 (1998), Nr. 4, S. 803–838

Misic u. Johnson 1999

MISIC, Mark M. ; JOHNSON, Kelsey L.: Benchmarking: a tool for Web site evaluation and improvement. In: *Internet Research: Electronic Networking Applications and Policy* 9 (1999), Nr. 5, S. 383–392

Mitchell u. a. 1997

MITCHELL, Ronald K. ; AGLE, Bradley R. ; WOOD, Donna J.: Toward a theorie of stakeholder identification and salience: Defining the principle of who and what really counts. In: *Academy of Management Review* 22 (1997), Nr. 4, S. 853–886

Mobasher u. a. 2002

MOBASHER, Bamshad ; DAI, Honghua ; LUO, Tao ; NAKAGAWA, Miki: Discovery and Evaluation of Aggregate Usage Profiles for Web Personalization. In: *Data Min. Knowl. Discov.* 6 (2002), Nr. 1, S. 61–82

Mobasher u. a. 1996

MOBASHER, Bamshad ; JAIN, Namit ; HAN, Eui-Hong (Sam) ; SRIVASTAVA, Jaideep: Web Mining: Pattern Discovery from World Wide Web Transactions / Department of Computer Science, University of Minnesota, Minneapolis. 1996 (TR-96050). – Forschungsbericht

Moe u. Fader 2004

MOE, Wendy W. ; FADER, Peter S.: Capturing Evolving Visit Behavior in Clickstream Data. In: *Journal of Interactive Marketing* 18 (2004), Nr. 1, S. 5–19

Molla u. Licker 2001

MOLLA, Alemayehu ; LICKER, Paul S.: e-Commerce Systems Success: An Attempt to Extend and Respecify the DeLone and MacLean Model of IS Success. In: *Journal of Electronic Commerce Research* 2 (2001), Nr. 4

Moon u. Kim 2001

MOON, Ji-Won ; KIM, Young-Gul: Extending the TAM for a World-Wide-Web context. In: *Information & Management* 38 (2001), Nr. 4, S. 217–230

Moustakis u. a. 2004

MOUSTAKIS, Vassilis S. ; LITOS, Charalambos ; DALIVIGAS, Andreas ; TSIRONIS, Loukas: Website Quality Assessment Criteria. In: *Proceedings of the Ninth International Conference on Information Quality (ICIQ-04)*, 2004, S. 59–73

Mukherjea u. a. 1995

MUKHERJEA, Sougata ; FOLEY, James D. ; HUDSON, Scott: Visualizing Complex Hypermedia Networks through Multiple Hierarchical Views. In: *Proceedings CHI'95*, 1995

Muylle u. a. 2004

MUYLLE, Steve ; MOENAERT, Rudy ; DESPONTIN, Marc: The conceptualization and empirical validation of web site user satisfaction. In: *Information & Management* 41 (2004), S. 543–560

Negash u. a. 2003

NEGASH, Solomon ; RYAN, Terry ; IGBARIA, Magid: Quality and effectiveness in Web-based customer support systems. In: *Information & Management* 40 (2003), S. 757–768

Netcraft.com 2008

June 2008 Web Server Survey. <http://news.netcraft.com/archives/2008/06/index.html>. Version: 2008. – Letzter Zugriff: 16.07.2008

Nielsen 1994

NIELSEN, Jakob: Heuristic Evaluation. In: NIELSEN, Jakob (Hrsg.) ; MACK, Robert L. (Hrsg.): *Usability Inspection Methods*. New York : John Wiley & Sons, 1994, Kapitel 2

Nielsen 2000

NIELSEN, Jakob: *Designing Web Usability*. Indianapolis : New Riders, 2000

Nielsen u. Mack 1994

NIELSEN, Jakob (Hrsg.) ; MACK, Robert L. (Hrsg.): *Usability Inspection Methods*. New York : John Wiley & Sons, 1994

Oates 2002

OATES, Briony J.: Foot and Mouth Disease: Lessons from local government websites. In: *2nd European Conference on e-Government ECEG 2002*. Oxford, UK, 2002, S. 341–352

Oliver 1999

OLIVER, Richard L.: Whence Consumer Loyalty? In: *Journal of Marketing* 63 (1999), S. 33–44

Olsina u. a. 1999

OLSINA, Luis ; GODOY, Daniela ; LAFUENTE, Guillermo ; ROSSI, Gustavo: Specifying Quality Characteristics and Attributes for Websites. In: *ICSE '99 Web Engineering Workshop*. Los Angeles, USA, 1999

Olsina u. Rossi 2002

OLSINA, Luis ; ROSSI, Gustavo: A Quantitative Method for Quality Evaluation of Web Sites and Applications. In: *IEEE Multimedia* 9 (2002), Nr. 4

Osei-Bryson u. Ko 2004

OSEI-BRYSON, Kweku-Muata ; KO, Myung: Applying Data Mining Techniques to Understand the Impact of Information Technology on Organizational Productivity. In: *Proceedings of the Tenth Americas Conference on Information Systems*. New York, NY, USA, 2004

Pagani 2006

PAGANI, Margherita: Determinants of adoption of High Speed Data Services in the business market: Evidence for a combined technology acceptance model with task technology fit model. In: *Information & Management* 43 (2006), S. 847–860

Palmer 2002

PALMER, Jonathan W.: Web Site Usability, Design, and Performance Metrics. In: *Information Systems Research* 13 (2002), Nr. 2

Parasuraman u. a. 1988

PARASURAMAN, A. ; ZEITHAML, Valerie A. ; BERRY, Leonard L.: SERVQUAL: A multi-item scale for measuring consumer perceptions of service quality. In: *Journal of Retailing* 61 (1988), Nr. 1, S. 12–40

Pardue u. Landry 2001

PARDUE, Harold ; LANDRY, Jeff: Evaluation of Alternative Interface Designs for E-Tail Shopping: An Empirical Study for Three Generalized Hierarchical Navigation Schemes. In: *Proceedings of the Seventh Americas Conference on Information Systems*. Boston, MA, USA, 2001

Paré 2004

PARÉ, Guy: Investigating Information Systems with Positivist Case Study. In: *Communications of the AIS* 13 (2004), S. 233–264

Pavlou 2001

PAVLOU, Paul A.: Integrating Trust in Electronic Commerce With the Technology Acceptance Model: Model Development and Validation. In: *Proceedings of the Seventh Americas Conference on Information Systems*. Boston, MA, USA, 2001

Pearce 2003

PEARCE, Liz: *Institutional Portals: A Review of Outputs.* <http://www.fair-portal.hull.ac.uk/downloads/iportaloutputs.pdf>. Version: 2003. – Letzter Zugriff: 10.11.2004

Pearce u. a. 2003

PEARCE, Liz ; CARPENTER, Leona ; MARTIN, Ruth: *Stakeholder Requirements for Institutional Portals.* www.fair-portal.hull.ac.uk/downloads/stakereq.pdf. Version: 2003. – Letzter Zugriff: 11.11.2004

Pirolli u. a. 1996

PIROLI, Peter ; PITKOW, J. ; RAO, R.: Silk from a sow's ear: extracting usable structures from the Web. In: *Conference proceedings on Human factors in computing systems*. Vancouver, British Columbia, Canada : ACM Press, 1996, S. 118–125

Pitt u. a. 1995

PITT, Leyland F. ; WATSON, Richard T. ; KAVAN, C. B.: Service Quality: A Measure of Information Systems Effectiveness. In: *MIS Quarterly* 19 (1995), Juni, Nr. 2, S. 173–187

Pohle u. Spiliopoulou 2002

POHLE, Carsten ; SPILIOPOULOU, Myra: Building and Exploiting Ad Hoc Concept Hierarchies for Web Log Analysis. In: *International Conference on Data Warehousing and Knowledge Discovery*, 2002, S. 83–93

Rai u. a. 2002

RAI, Arun ; LANG, Sandra S. ; WELKER, Robert B.: Assessing the Validity of IS Success Models: An Empirical Test and Theoretical Analysis. In: *Information Systems Research* 13 (2002), Nr. 1, S. 50–69

Rässler u. Riphahn 2006

RÄSSLER, Susanne ; RIPHAHN, Regina: Survey item nonresponse and its treatment. In: *AStA Advances in Statistical Analysis* 90 (2006), March, Nr. 1, S. 217–232

Reicheld u. Shefter 2000

REICHELD, Frederick F. ; SHEFTER, Phil: E-Loyalty: Your Secret Weapon on the Web. In: *Harvard Business Review* (2000), July-August

Ridings u. a. 2002

RIDINGS, Catherine M. ; GEFEN, David ; ARINZE, Bay: Some antecedents and effects of trust in virtual communities. In: *Journal of Strategic Information Systems* 11 (2002), S. 271–295

van Riel u. a. 2001

RIEL, Allard C. R. ; LILJANDER, Veronica ; JURRIËNS, Petra: Exploring consumer evaluations of e-services: a portal site. In: *International Journal of Service Industry Management* 12 (2001), Nr. 4, S. 359–377

Ringle u. a. 2007

RINGLE, C. M. ; WENDE, S. ; ALEXANDER, W.: *SmartPLS 2.0 M3*. Universität Hamburg. www.smartpls.de. Version: 2007. – Letzter Zugriff: 21.07.2007

Rossi u. a. 2004

ROSSI, Peter H. ; LIPSEY, Mark W. ; FREEMAN, Howard E.: *Evaluation: A Systematic Approach*. 5. Thousand Oaks, USA : Sage Publications, 2004

Rubin 1987

RUBIN, Donald B.: *Multiple Imputation for Nonresponse in Surveys*. New York : Wiley, 1987

Rubin 1996

RUBIN, Donald B.: Multiple Imputation After 18+ Years. In: *Journal of the American Statistical Association* 91 (1996), Nr. 434

Sampson u. Manouselis 2005

SAMPSON, Demetrios ; MANOUSELIS, Nikos: A Flexible Evaluation Framework for Web Portals Based on Multi-Criteria Analysis. In: TATNALL, Arthur (Hrsg.): *Web*

Portals: The New Gateways to Internet Services. London : Idea Group Publishing, 2005, S. 185–210

Schafer 1997

SCHAFER, Joseph L.: *Analysis of Incomplete Multivariate Data.* London : Chapman & Hall, 1997

Schneeweiss 1991

SCHNEEWEISS, Christoph: *Planung 1, Systemanalytische und entscheidungstheoretische Grundlagen.* Bd. 1. Berlin : Springer, 1991

SearchCIO-Midmarket.com 2003

Midmarket CIO Definitions - Portal. http://searchcio-midmarket.techtarget.com/sDefinition/0,,sid183_gci212810,00.html. Version: 2003. – Letzter Zugriff: 16.07.2008

Searle 2005

SEARLE, Ian: Portals in Large Enterprises. In: TATNALL, Arthur (Hrsg.): *Web Portals: The New Gateways to Internet Services.* London : Idea Group Publishing, 2005, S. 119–171

Seddon 1997

SEDDON, Peter B.: A Respecification and Extension of the DeLone and McLean Model of IS Success. In: *Information Systems Research* 8 (1997), Nr. 3, S. 240–253

Seddon u. a. 1999

SEDDON, Peter B. ; STAPLES, Sandy ; PATNAYAKUNI, Ravi ; BOWTELL, Matthew: Dimensions of Information Systems Success. In: *Communications of the AIS* 2 (1999)

Sedera u. Tan 2005

SEDERA, Darshana ; TAN, Felix Ter C.: User Satisfaction: An Overarching Measure of Enterprise System Success. In: *Proceedings of the Pacific Asia Conference on Information Systems*, 2005

Shadish u. a. 1991

SHADISH, William R. ; COOK, Thomas D. ; LEVITON, Laura C.: *Foundations of Program Evaluation: Theories of Practice.* London : SAGE Publications, 1991

Shannon u. Weaver 1949

SHANNON, Claude E. ; WEAVER, Warren: *The Mathematical Theorie of Communication.* Urbana, IL : University of Illinois, 1949

Shih 2004

SHIH, Hung-Pin: An empirical study on predicting user acceptance of e-shopping on the Web. In: *Information & Management* 41 (2004), S. 351–368

Sieber u. Valor-Sabatier 2005

SIEBER, Sandra ; VALOR-SABATIER, Josep: Competitive Dynamics of General Portals. In: TATNALL, Arthur (Hrsg.): *Web Portals: The New Gateways to Internet Services.* London : Idea Group Publishing, 2005, S. 64–79

Song u. Zahedi 2001

SONG, Jaeki ; ZAHEDI, Fatemeh: Web Design in E-Commerce: A Theory and Empirical Analysis. In: *Twenty-Second International Conference on Information Systems*, 2001

Speier u. Poston 2001

SPEIER, Cheri ; POSTON, Robin: Web Site Acceptance: The Effects of Task Type. In: *Proceedings of the Seventh Americas Conference on Information Systems*. Boston, MA, USA, 2001

Spertus 1997

SPERTUS, Ellen: ParaSite: mining structural information on the Web. In: *Computer Networks and ISDN Systems* 29 (1997), S. 1205–1215

Spiliopoulou 2000

SPILIOPOULOU, Myra: Web Usage Mining for Web Site Evaluation. In: *Communications of the ACM* 43 (2000), Nr. 8

Spiliopoulou u. Berendt 2001

SPILIOPOULOU, Myra ; BERENDT, Bettina: Kontrolle der Präsentation und Vermarktung von Gütern im WWW anhand von Data-Mining-Techniken. In: HIPPNER, H. (Hrsg.) ; KÜSTERS, U. (Hrsg.) ; MEYER, M. (Hrsg.) ; WILDE, K.D. (Hrsg.): *Handbuch Data Mining im Marketing - Knowledge Discovery in Databases*. Wiesbaden, Germany : Vieweg, 2001, S. 855–873

Spiliopoulou u. Berendt 2002

SPILIOPOULOU, Myra ; BERENDT, Bettina: Wie werden Surfer zu Kunden? – Navigationsanalyse zur Ermittlung des Konversionspotenzials verschiedener Sitebereiche. In: HIPPNER, Hajo (Hrsg.) ; MERZENICH, Melanie (Hrsg.) ; WILDE, Klaus (Hrsg.): *Handbuch Web Mining im Marketing. Konzepte, Systeme, Fallstudien* Bd. 1. Vieweg, 2002, Kapitel 3.5

Spiliopoulou u. Faulstich 1998

SPILIOPOULOU, Myra ; FAULSTICH, Lukas C.: WUM: A Web Utilization Miner. In: *Workshop on the Web and Data Bases (WebDB98)*, 1998, S. 109–115

Spiliopoulou u. a. 2003

SPILIOPOULOU, Myra ; MOBASHER, Bamshad ; BERENDT, Bettina ; NAKAGAWA, Niki: A Framework for the Evaluation of Session Reconstruction Heuristics in Web Usage Analysis. In: *INFORMS Journal of Computing, Special Issue on Mining Web-Based Data for E-Business Applications* 15 (2003), Nr. 2

Spiliopoulou u. Pohle 2001

SPILIOPOULOU, Myra ; POHLE, Carsten: Data Mining for Measuring and Improving the Success of Web Sites. In: *Journal of Data Mining and Knowledge Discovery* 5 (2001), Nr. 1, S. 85–114

Spiliopoulou u. a. 2002

SPILIOPOULOU, Myra ; POHLE, Carsten ; TELTZROW, Maximilian: Modelling Web Site Usage with Sequences of Goal-Oriented Tasks. In: WEINHARDT, Christof (Hrsg.) ; HOLTMANN, Carsten (Hrsg.): *Proc. of Workshop E-Commerce: Netze, Märkte, Technologien, in Multikonferenz der Wirtschaftsinformatik*, Physica-Verlag, 2002, S. 203–221

Srikant u. Agrawal 1996

SRIKANT, Ramakrishnan ; AGRAWAL, Rakesh: Mining sequential patterns: Generalizations and performance improving. In: *Proceedings of EDBT*. Avignon, France, 1996

Srivastava u. a. 2000

SRIVASTAVA, Jaideep ; COOLEY, Robert ; DESHPANDE, Mukund ; TAN, Pang-Ning: Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data. In: *SIGKDD Explorations* 1 (2000), Nr. 2, S. 12–23

Staples u. Seddon 2004

STAPLES, Sandy ; SEDDON, Peter: Testing the technology-to-performance chain model. In: *Journal of Organizational and End User Computing* 16 (2004), Nr. 4, S. 17–36

Suchmann 1995

SUCHMANN, M. C.: Managing legitimacy: Strategic and institutional approaches. In: *Academy of Management Review* 20 (1995), S. 571–610

Sullivan 1997

SULLIVAN, Terry: Reading reader reaction: A proposal for inferential analysis of web server log files. In: *Proceedings of the Human Factors and the Web 3 Conference, Practices & Reflections*, 1997

Tan u. Kumar 2000

TAN, Pang-Ning ; KUMAR, Vipin: Modeling of Web Robot Navigational Patterns. In: *Proc. ACM WebKDD Workshop*, 2000

Tan u. Kumar 2002

TAN, Pang-Ning ; KUMAR, Vipin: Discovery of Web Robot Sessions based on their Navigational Patterns. In: *Data Mining and Knowledge Discovery* 6 (2002), Nr. 1, S. 9–35

Tatnall 2005

TATNALL, Arthur: Portals, Portals Everywhere. In: TATNALL, Arthur (Hrsg.): *Web Portals: The New Gateways to Internet Services*. London : Idea Group Publishing, 2005, S. 1–14

Tennehaus u. a. 2005

TENNEHAUS, M. ; VINZI, V. E. ; CHATELIN, Y.-M. ; LAURO, C.: PLS Path Modeling. In: *Computational Statistics & Data Analysis* 48 (2005), S. 159–205

Torkzadeh u. Dhillon 2002

TORKZADEH, Gholamreza ; DHILLON, Gurpreet: Measuring Factors that Influence the Success of Internet Commerce. In: *Information Systems Research* 13 (2002), Nr. 2, S. 187–204

Urbaczewski u. a. 2002

URBACZEWSKI, Andrew ; JESSUP, Leonard M. ; WHEELER, Bradley: Electronic Commerce Research: A Taxonomy and Synthesis. In: *Journal of Organizational Computing and Electronic Commerce* 12 (2002), Nr. 4, S. 263–305

Van Dyke u. a. 1995

VAN DYKE, Thomas P. ; KAPPELMAN, Leon A. ; PRYBUTOK, Victor R.: Measuring Information Systems Service Quality: Concerns on the Use of the SERVQUAL Questionnaire. In: *MIS Quarterly* (1995), Juni, S. 173–187

Vatanasakdakul u. D'Ambrä 2007

VATANASAKDAKUL, Savanid ; D'AMBRA, John: A conceptual model for e-commerce adoption in developing countries: a task-technology fit perspective. In: *International Journal of Information Technology & Management* 6 (2007), Nr. 2–4

Venkatesh 2000

VENKATESH, Viswanath: Determinants of Perceived Ease of Use: Integrating Control, Intrinsic Motivation and Emotion into the Technology Acceptance Model. In: *Information Systems Research* 11 (2000), Nr. 4, S. 342–365

Venkatesh u. Davis 2000

VENKATESH, Viswanath ; DAVIS, Fred D.: A Theoretical Extension of the Technology Acceptance Modell: Four Longitudinal Studies. In: *Management Science* 46 (2000), Nr. 2, S. 186–204

Venkatesh u. a. 2003

VENKATESH, Viswanath ; MORRIS, Michael G. ; DAVIS, Gordon B. ; DAVIS, Fred D.: User Acceptance of Information Technology: Toward a Unified View. In: *MIS Quarterly* 27 (2003), Nr. 3, S. 425–478

W3C 1995

W3C: *Logging Control In W3C httpd*. www.w3.org/Daemon/User/Config/Logging.html. Version: 1995. – Letzter Zugriff: 18.11.2006

Wang u. Strong 1996

WANG, Richard Y. ; STRONG, Diane M.: Beyond Accuracy: What Data Quality Means to Data Consumers. In: *Journal of Management Information Systems* 12 (1996), Nr. 4, S. 5–34

Wang u. Butler 2003

WANG, Xiaoqing ; BUTLER, Brian S.: Individual Technology Acceptance under Conditions of Change. In: *Twenty-Fourth International Conference on Information Systems*, 2003

Wang u. Tang 2003

WANG, Yi-Shun ; TANG, Tzung-I: Assessing Customer Perceptions of Website Service Quality in Digital Marketing Environments. In: *International Journal of End User Computing* 15 (2003), Nr. 3, S. 14–31

Wang u. a. 2001

WANG, Yi-Shun ; TANG, Tzung-I ; TANG, Jeung-tai E.: An instrument for measuring customer satisfaction toward Web sites that market digital products and services. In: *Journal of Electronic Commerce Research* 2 (2001), Nr. 3, S. 89–102

Warrington u. a. 2000

WARRINGTON, Traci B. ; ABGRAB, Nadia J. ; CALDWELL, Helen M.: Building trust to develop competitive advantage in e-business relationships. In: *Competitiveness Review* 10 (2000), Nr. 2, S. 160–168

Watson u. a. 1998

WATSON, Richard T. ; PITT, Leyland F. ; KAVAN, C. B.: Measuring Informations Systems Service Quality: Lessons from two Longitudinal Case Studies. In: *MIS Quarterly* 22 (1998), S. 61–79

Webb u. Webb 2002

WEBB, Harold W. ; WEBB, Linda A.: B2C Electronic Commerce Websites: An Analysis of Quality Factors. In: *Proceedings of the Eighth Americas Conference on Information Systems*. Dallas, TX, USA, 2002, S. 340–347

Weber 1967

WEBER, Max: Über einige Kategorien der verstehenden Soziologie. Version: 1967. www.uni-potsdam.de/u/paed/pia/index.htm. In: WINCKELMANN, Johannes (Hrsg.): *Gesammelte Aufsätze zur Wissenschaftslehre*, Kapitel 8, 427–474. – Online-Ressource. – Letzter Zugriff: 28.12.2007

Wikipedia.org 2008a

Entscheidung. <http://de.wikipedia.org/wiki/Entscheidung>. Version: 2008. – Letzter Zugriff: 16.07.2008

Wikipedia.org 2008b

Information systems. http://en.wikipedia.org/wiki/Information_system. Version: 2008. – Letzter Zugriff: 16.07.2008

Wiktionary.org 2008

Intrinsisch. <http://de.wiktionary.org/wiki/intrinsisch>. Version: 2008. – Letzter Zugriff: 16.07.2008

Wille 1992

WILLE, Rudolf: Concept lattices and conceptual knowledge systems. In: *Computers & Mathematics with Applications* 23 (1992), Nr. 6–9, S. 493–515

Witten u. Frank 2005

WITTEN, Ian H. ; FRANK, Eibe: *Data Mining: Practical Machine Learning Tools and Techniques*. 2. San Francisco : Morgan Kaufmann, 2005

Wojtkowski u. Major 2005

WOJTKOWSKI, Wita ; MAJOR, Marshall: On Portals: A Parsimonious Approach. In: TATNALL, Arthur (Hrsg.): *Web Portals: The New Gateways to Internet Services*. London : Idea Group Publishing, 2005, S. 15–39

Woodrooff u. Burg 2003

WOODROOF, Jon ; BURG, William: Satisfaction/dissatisfaction: are users predisposed? In: *Information & Management* 40 (2003), S. 317–324

Xiao u. Dasgupta 2002

XIAO, Li ; DASGUPTA, Subhasish: Measurement of User Satisfaction with Web-based Information Systems: An Empirical Study. In: *Proceedings of the Eighth Americas Conference on Information Systems*. Dallas, TX, USA, 2002

Yin 2003

YIN, Robert K.: *Case Study Research: Design and Methods*. 3. London : Sage, 2003

Yoo u. Donthu 2001

YOO, Boonghee ; DONTHU, Naveen: Developing a Scale to Measure the Perceived Quality of an Internet Shopping Site (SITEQUAL). In: *Quarterly Journal of Electronic Commerce* 2 (2001), Nr. 1, S. 31–46

Yuthas u. Young 1998

YUTHAS, Kristi ; YOUNG, Scott: Material matters: Assessing the effectiveness of materials management IS. In: *Information & Management* 33 (1998), S. 115–124

Zaïane u. a. 1998

ZAÏANE, Osmar ; XIN, Man ; HAN, Jiawei: Discovering web access patterns and trends by applying OLAP and data mining technology on web logs. In: *Advances of Digital Libraries*. Santa Barbara, CA, 1998, S. 19–29

Zamir u. a. 1997

ZAMIR, Oren ; ETZIONI, Oren ; MADANI, Omid ; KARP, Richard M.: Fast and Intuitive Clustering of Web Documents. In: *Proceedings of the Third International Conference on Knowledge Discovery and Data Mining*. CA, USA : AAAI Press, 1997, S. 287–290

Zeithaml u. a. 2002

ZEITHAML, Valarie A. ; PARASURAMAN, A. ; MALHOTRA, A.: Service Quality Delivery Through Web Sites: A Critical Review of Extant Knowledge. In: *Journal of the Academy of Marketing Science*. 30 (2002), Nr. 4, S. 362–375

Zeithaml u. a. 1996

ZEITHAML, Valarie A. ; BERRY, Leonard L. ; PARASURAMAN, A.: The Behavioral Consequences of Service Quality. In: *Journal of Marketing* 60 (1996), S. 31–46

Zhang u. von Dran 2001

ZHANG, Ping ; DRAN, Gisela M.: User Expectations and Rankings of Quality Factors in Different Web Site Domains. In: *International Journal of Electronic Commerce / Winter 2001–2002* 6 (2001), Nr. 2, S. 9–33

Zviran u. Erlich 2003

ZVIRAN, Moshe ; ERLICH, Zippy: Measuring IS User Satisfaction: Review and Implications. In: *Communications of the AIS* 12 (2003)

Anhang A

Ergänzungsansätze bei unvollständigen Daten

Im vorliegenden Anhang wird eine kurze Einführung zu Ergänzungsansätzen bei unvollständigen Daten gegeben. Ferner werden die Prozeduraufälle für die hier zur multiplen Imputation verwendete Software SAS 9.1.3 beschrieben.

A.1 Fehlmechanismen in Daten

Sei $Y = (y_{ij})$ eine $n \times K$ Matrix vollständiger aber nicht vollständig beobachtbarer Daten mit y_{ij} als dem Wert des Attributes Y_j für Datensatz i , mit $i = 1, \dots, n$ und $j = 1, \dots, K$. Daraus lässt sich eine Indikatormatrix $M = (m_{ij})$ für nicht beobachtete Werte ableiten, mit $m_{ij} = 1$ falls y_{ij} nicht beobachtet wurde und $m_{ij} = 0$ sonst. Der Typ fehlender Daten ist dann charakterisiert durch die bedingte Verteilung von M gegeben Y , notiert als $f(M|Y, \phi)$, mit ϕ als weiteren unbekannten Parametern.

Es lassen sich drei Typen von Fehlmechanismen in Daten unterscheiden (Little u. Rubin 2004, S. 11 ff.):

- *Missing completely at random* (MCAR),
- *Missing at random* (MAR),
- *Not missing at random* (NMAR).

MCAR liegt vor, wenn die nicht beobachteten Werte in Y unabhängig von den beobachteten sind, d.h. $f(M|Y, \phi) = f(M|\phi)$ für alle Y, ϕ . Ein solcher Fall ergäbe sich beispielsweise, wenn ein Umfrageteilnehmer durch Münzwurf entscheiden würde, ob er eine Frage beantwortet oder nicht.

MAR basiert auf einer weniger restriktiven Annahme. Sei Y_{obs} der beobachtbare und Y_{mis} der nicht beobachtbare Teil von Y . Nicht beobachtete Werte hängen im Fall MAR von Y_{obs} ab, jedoch nicht von Y_{mis} , d.h. $f(M|Y, \phi) = f(M|Y_{obs}, \phi)$

für alle Y_{obs}, ϕ . Aus der Beantwortung anderer Fragen kann in diesem Fall die Antwort für eine nicht beantwortete Frage im Fragebogen abgeleitet werden. Um diese Zusammenhänge erfassen zu können, muss ein gewisser Teil der Befragten diese Frage dafür allerdings beantwortet haben.

NMAR ist gegeben, wenn die bedingte Verteilung von M nur von Y_{mis} abhängt. In diesem Fall lässt sich $f(M|Y, \phi)$ nicht vereinfachen. Beispiel hierfür wären sehr gut verdienende Umfrageteilnehmer, die ihr Einkommen nicht angeben und bei denen das Einkommen auch nicht aus der Beantwortung der anderen Fragen ableitbar wäre.

A.2 Ergänzungsmethoden

In der Praxis werden unvollständige Daten bei einer statistischen Auswertung in der Regel ignoriert, d. h. nur die vollständigen Fälle werden betrachtet (engl.: *complete case analysis*). Diese Vorgehensweise impliziert MCAR als Fehlmechanismus. In den meisten Fällen kann man jedoch davon ausgehen, dass MAR vorliegt (siehe Beispiele in Schafer 1997, S. 20 ff.) und eine Analyse auf Basis der vollständigen Daten dann zu systematisch mit Fehlern behafteten Ergebnissen (engl.: *bias*) führen würde (King u. a. 2001; Schafer 1997, S. 23 f.). Die meisten Methoden zur Behandlung fehlender Werte basieren daher auf der MAR-Annahme. Aber auch für den Fall, dass MCAR vorliegt, führt eine Imputation in der Regel zu valideren Ergebnissen, als die Analyse nur der vollständigen Daten (King u. a. 2001). Nur im Fall NMAR können beide Ansätze zu fehlerhaften Ergebnissen führen. Das vorliegen von NMAR kann jedoch per Definition (s. o.) anhand der Erhebungsdaten nicht nachgewiesen werden.

Vorhandene Imputationsansätze lassen sich in Methoden der einfachen (engl.: *single*) und der multiplen (engl.: *multiple*) Imputation gliedern. Bei der einfachen Imputation wird für jeden fehlenden Wert genau ein Wert eingesetzt.¹ Vorteil dieses Ansatzes ist, dass die Daten im Anschluss sofort mit Standardmethoden zur statistischen Auswertung analysiert werden können. Außerdem kann je nach verwendeter Methode internes Wissen des Analysten in die Datenvervollständigung einfließen. Der Nachteil ist, dass fehlende Werte behandelt werden, als wären sie bekannt. Daraus ergibt sich eine künstliche Verringerung der Varianz bei diesen Werten sowie ein Bias bei aus diesen Werten errechneten Statistiken. Multiple Imputation dagegen beinhaltet die Vorteile der einfachen Imputation ohne eben aufgeführte Nachteile. Hierbei werden m Werte für einen fehlenden Wert eingesetzt, d. h. durch den Imputationsvorgang werden aus der ursprünglichen Datenmenge mit fehlenden Werten m vollständige Datenmengen erzeugt. Diese werden jeweils separat analysiert und die daraus abgeleiteten m Schätzer für eine Statistik (z. B. Mittelwert

¹Methoden der einfachen Imputation sind z. B. (1) einsetzen von Mittelwert oder Median, (2) herleiten von Werten durch Vergleich mit ähnlichen Datensätzen (a) innerhalb der betrachteten Datentabelle (engl.: *hot-deck imputation*) oder (b) aus anderen Datentabellen (engl.: *cold-deck imputation*).

oder Varianz) werden zusammengeführt. Auf diese Weise wird im Gegensatz zur einfachen Imputation die Unsicherheit bezüglich der einzusetzenden Werte, unter Bezug auf das zugrunde gelegte Modell der Daten, in Form einer höheren Varianz berücksichtigt.

Im Vergleich zur einfachen Imputation hat die multiple Imputation drei Nachteile: (1) Der Imputationsaufwand ist größer, da die Imputation m -mal durchgeführt wird; (2) mehr Speicherkapazität wird dadurch benötigt; (3) der Analyseaufwand ist größer, da m Datentabellen analysiert und die Ergebnisse zusammengeführt werden müssen (Rubin 1987, S. 11 ff.).

Untersuchungen haben gezeigt, dass die Multiple Imputation, selbst unter einem naiven Modell, zu besseren Ergebnissen führt, als die einfache (Heitjan u. Rubin 1990). Da die Korrektheit der Ergebnisse und nicht der Analyseaufwand als Argument für die Wahl eines Ansatzes ausschlaggebend sein sollte, wäre die Multiple Imputation daher immer vorzuziehen.

Inwieweit diese zu validen Aussagen führt, hängt von der jeweils gewählten Methode und den damit einhergehenden Modellannahmen zu den Daten ab. Rubin (1987, S. 118 f.) nennt in diesem Zusammenhang den Begriff *proper*, welcher hier mit "geeignet" übersetzt werden kann. Eine geeignete multiple Imputationsmethode führt dazu, dass aus den imputierten Daten dieselben Schlussfolgerungen gezogen werden können, wie sie auf Basis der vollständigen Daten ableitbar wären.

Die Eignung einer Methode auf Basis der beobachteten Daten nachzuweisen, ist allerdings sehr schwierig. Rubin (1987, S. 125 ff.) schlägt daher den Einsatz von Bayes-Verfahren vor, die unter einem korrekt angenommenen Typ des Fehlmechanismus und einem zu den Daten passenden Modell in „großen“ Stichproben eine geeignete multiple Imputation ermöglicht. Hierbei ist es wichtig, dass alle verfügbaren Variablen in den Imputationsvorgang einbezogen werden. Variablen, die bei der Imputation nicht einbezogen wurden, wohl aber in der statistischen Analyse betrachtet werden, können zu einem Bias in den Ergebnissen führen. Andersherum ist der mögliche Verlust an Präzision durch Einbezug von für die statistische Analyse unwichtigen Variablen ein geringer zu zahlender Preis für die dadurch verbesserte Validität der Imputation (Rubin 1996).

Liegt eine geeignete Methode vor und ist der Anteil an fehlenden Werten nicht übermäßig groß (z. B. unter 30%), reicht in der Regel schon eine geringe Anzahl an Imputationen ($m \leq 5$) für eine valide Einschätzung der fehlenden Werte aus (Rubin 1987, S. 113 ff.).

Die Ermittlung von Schätzwerten basiert auf folgenden Annahmen (Schafer 1997, S. 108):

Sei Q eine Statistik (Mittelwert, Varianz usw.) der vollständigen Daten Y und sei \hat{Q} der auf den erhobenen, unvollständigen Daten basierende Schätzer für Q . Des Weiteren sei \hat{U} der Varianzschatzer von \hat{Q} ($\hat{U} = \widehat{\text{var}}(\hat{Q})$). \hat{Q} und \hat{U} sind somit Schätzfunktionen auf Y , mit $\hat{Q}(Y) = \hat{Q}(Y_{\text{obs}}, Y_{\text{mis}})$ und $\hat{U}(Y) = \hat{U}(Y_{\text{obs}}, Y_{\text{mis}})$. Der Multiplen Imputation liegt die Annahme zugrunde, dass diese Schätzfunktionen Näherungen erster Ordnung der A-posteriori-Verteilung von Q sind, mit

$$\hat{Q}(Y_{obs}, Y_{mis}) \approx E(Q|Y_{obs}, Y_{mis}) \quad (\text{A.1})$$

und

$$\hat{U}(Y_{obs}, Y_{mis}) \approx E(V|Y_{obs}, Y_{mis}) \quad (\text{A.2})$$

bei geeignetem zugrunde gelegtem Modell der vollständigen Daten und der A-priori-Verteilung. Ferner wird angenommen, dass für die Schätzfunktionen, falls unbekannt, zumindest approximativ ($n \rightarrow \infty$) die Normalverteilung gilt:

$$\frac{Q - \hat{Q}}{\sqrt{\hat{U}}} \sim N(0, 1). \quad (\text{A.3})$$

Bei Bedarf sollten die Schätzfunktionen geeignet transformiert werden, um diese asymptotische Normalverteiltheit zu gewährleisten.

Die Zusammenführung der m abgeleiteten Schätzer erfolgt nach einem simplen Prinzip (Schafer 1997, S. 107 ff. in Anlehnung an Rubin 1987, S. 75 ff.):

Die Zusammenführung der m Punktschätzer für Q erfolgt durch Bildung des Mittelwertes \bar{Q} über alle $\hat{Q}^{(t)}$ für alle $t = 1, 2, \dots, m$ mit

$$\bar{Q} = \frac{1}{m} \sum_{t=1}^m \hat{Q}^{(t)}. \quad (\text{A.4})$$

Die Zusammenführung der m Varianzschätzer für \bar{Q} besteht aus der Komponente \bar{U} , d. h. dem Mittelwert der m Varianzschätzer (engl.: *within-imputation variance*), mit

$$\bar{U} = \frac{1}{m} \sum_{t=1}^m \hat{U}^{(t)} \quad (\text{A.5})$$

und der Komponente \hat{B} , d. h. der Varianz zwischen den m Punktschätzern $\hat{Q}^{(t)}$ (engl.: *between-imputation variance*), mit

$$\hat{B} = \frac{1}{m-1} \sum_{t=1}^m (\hat{Q}^{(t)} - \bar{Q})^2. \quad (\text{A.6})$$

Die Zusammenführung dieser zwei Komponenten ergibt die totale Varianz \hat{T} der Imputation für den betrachteten Punktschätzer \hat{Q} mit

$$\hat{T} = \bar{U} + \left(\frac{1}{m-1} \right) \hat{B}. \quad (\text{A.7})$$

Hypothesentests können dann auf Basis der t-Verteilung durchgeführt werden mit

$$\frac{Q - \hat{Q}}{\sqrt{\hat{T}}} \sim t_\nu. \quad (\text{A.8})$$

Die Anzahl der Freiheitsgrade ν errechnet sich aus

$$\nu = (m - 1) \left[1 + \frac{\bar{U}}{(1 + \frac{1}{m})\hat{B}} \right]^2. \quad (\text{A.9})$$

Zwei Indikatoren zur Einschätzung des Einflusses fehlender Werte auf den Grad der Unsicherheit von Auswertungsergebnissen sind r und $\hat{\lambda}$ mit

$$r = \frac{(1 + \frac{1}{m})\hat{B}}{\bar{U}} \quad (\text{A.10})$$

als den relativen Anstieg in der Varianz durch die Imputation und

$$\hat{\lambda} = \frac{(r + 2)/(\nu + 3)}{r + 1} \quad (\text{A.11})$$

als einen Schätzer für den Anteil der fehlenden Informationen über Q .

Die relative Effizienz einer finiten Anzahl von m Imputationen im Vergleich zu $m \rightarrow \infty$ lässt sich approximativ als eine Funktion von m und $\hat{\lambda}$ berechnen (Rubin 1987, S. 114) mit

$$RE = \left(1 + \frac{\hat{\lambda}}{m} \right)^{-1}. \quad (\text{A.12})$$

A.3 Imputation mit SAS (PROC MI)

Die in dieser Arbeit verwendete Software zur Imputation fehlender Werte ist SAS 9.1.3. Die verwendete Prozedur ist PROC MI. Wie in Abschnitt 6.3 beschrieben, werden hierfür zwei Funktionen von PROC MI genutzt. Zuerst wird mittels einer MCMC-Simulation das Muster der fehlenden Werte in der Datentabelle monotonisiert (siehe Beispiel in Tabelle 6.13 auf Seite 119). Im Anschluss werden die verbleibenden fehlenden Werte auf Basis logistischer Regression ergänzt.

Für die Monotonisierung der Fehlmuster wird die Funktion MCMC mit der Option MONOTONE aufgerufen (Abbildung A.1). Hierfür muss eine feste Reihenfolge der Variablen mit der Option VAR vorgegeben werden. Da der MCMC-Algorithmus mit kontinuierlichen Werten arbeitet, wird zur Imputation „kategorischer“ Werte mit der Option ROUND=1 eine Auf- bzw. Abrundung der imputierten Werte auf den nächsten Integer-Wert vorgegeben. Die Optionen MAXIMUM und MINIMUM geben die Intervallgrenzen für die zu imputierenden Werte an.

Im Anschluss an die Monotonisierung kann das für diese Daten besser geeignete Imputationsverfahren der logistischen Regression eingesetzt werden. Die entsprechende Funktion ist MONOTONE LOGISTIC (Abbildung A.2). Auch hier ist die Reihenfolge der Variablen durch die Option VAR vorzugeben. Außerdem sind die

```
proc mi data=in.dat maximum=5 minimum=1 round=1 seed=123456 out=outmon.dat;
mcmc impute=monotone;
var PORT INFQ1 INFQ2 INFQ3 INFQ4 INFQ5 SYSQ1 SYSQ2 SYSQ3 SYSQ4 SYSQ5 SYSQ6
ZUFR1 ZUFR2 WISS1 WISS2 LOYA1 LOYA2 VERT1 VERT2 RISK1 RISK2;
run;
```

Abbildung A.1: Aufruf der Prozedur PROC MI zur Monotonisierung der Muster der fehlenden Werte in der Datentabelle mittels MCMC-Simulation

Klassenvariablen mittels Option CLASS vorzugeben. Da im Prozeduraufruf keine Abhängigkeitsregeln für eine Klassenvariable vorgegeben sind, werden für die Imputation einer Variablen alle anderen Variablen als so genannte Kovariate berücksichtigt. Da bereits mit der MCMC-Funktion *m* Imputationsdatensätze erstellt wurden, wird mit der Option NIMPUTE=1 ein erneutes Vervielfachen der Datentabellen verhindert.

```
proc mi data=outmon.dat nimpute=1 seed=123456 out=out.dat;
class PORT INFQ1 INFQ2 INFQ3 INFQ4 INFQ5 SYSQ1 SYSQ2 SYSQ3 SYSQ4 SYSQ5 SYSQ6
ZUFR1 ZUFR2 WISS1 WISS2 LOYA1 LOYA2 VERT1 VERT2 RISK1 RISK2;
monotone logistic;
var PORT INFQ1 INFQ2 INFQ3 INFQ4 INFQ5 SYSQ1 SYSQ2 SYSQ3 SYSQ4 SYSQ5 SYSQ6
ZUFR1 ZUFR2 WISS1 WISS2 LOYA1 LOYA2 VERT1 VERT2 RISK1 RISK2;
by _Imputation_;
run;
```

Abbildung A.2: Aufruf der Prozedur PROC MI zur Imputation fehlender Werte in der Datentabelle mittels logistischer Regression

Anhang B

Histogramme zu den Umfrageergebnissen

Im vorliegenden Anhang sind die Erhebungsdaten zu den Portalumfragen in ihrer Originalform (vor der Imputation fehlender Werte) in Histogrammen dargestellt (Abbildungen B.1–B.4).

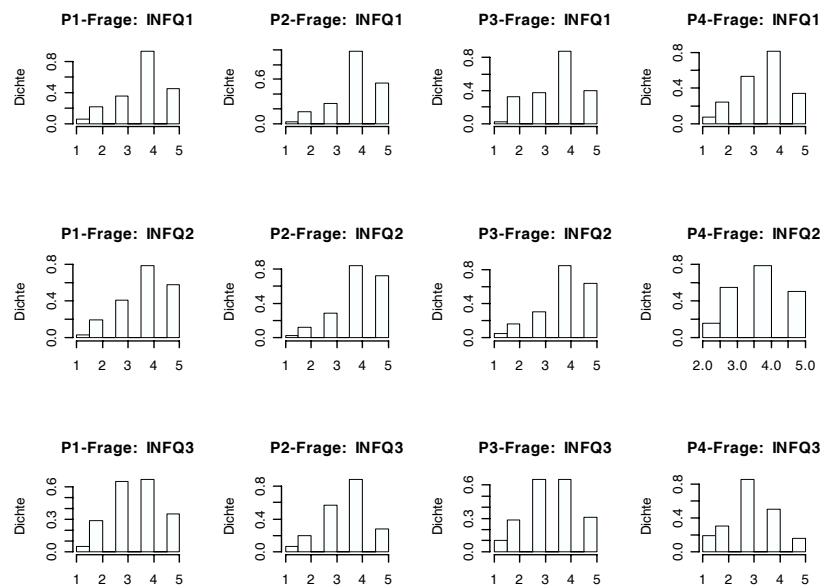


Abbildung B.1: Histogramme zu den erhobenen Daten der Portale P1–P4 (1)

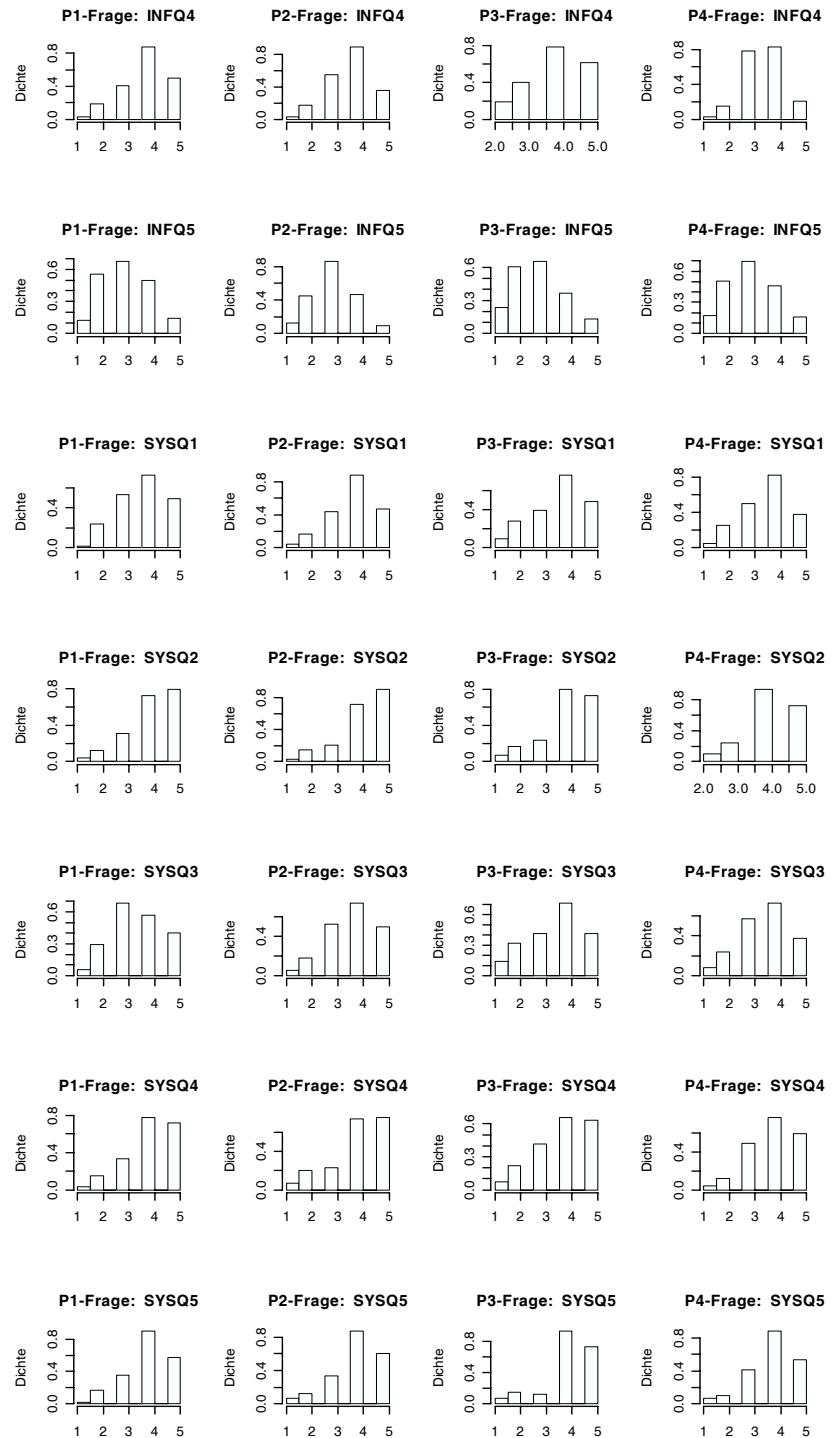


Abbildung B.2: Histogramme zu den erhobenen Daten der Portale P1–P4 (2)

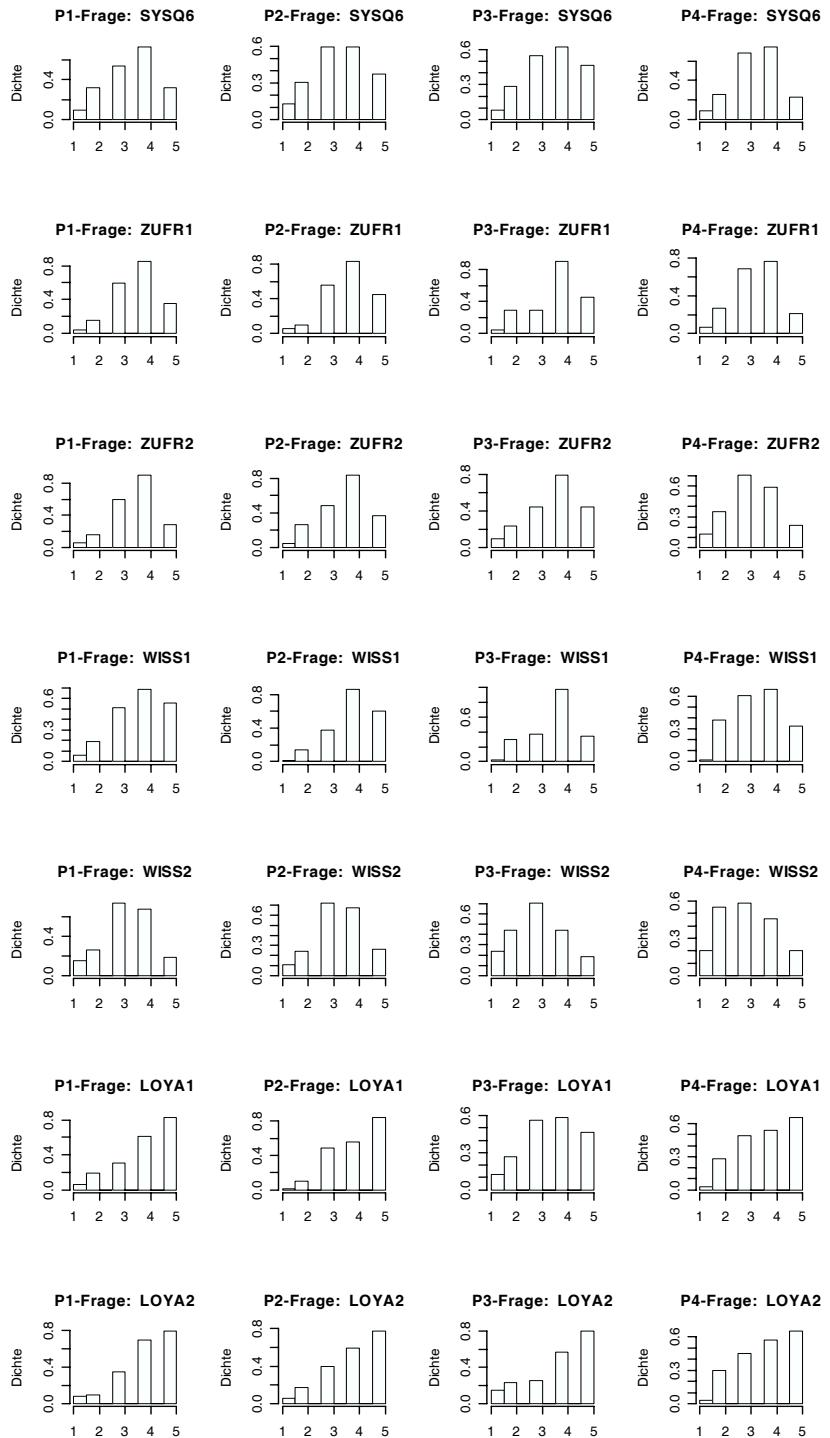


Abbildung B.3: Histogramme zu den erhobenen Daten der Portale P1–P4 (3)

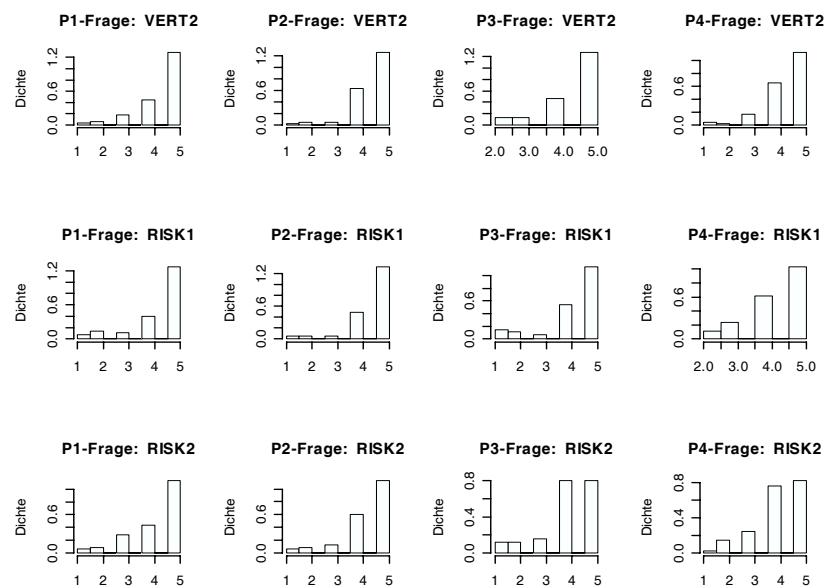


Abbildung B.4: Histogramme zu den erhobenen Daten der Portale P1–P4 (4)

Anhang C

Validierungsergebnisse zum Pfadmodell

Im vorliegenden Anhang sind die Validierungsergebnisse zum postulierten Pfadmodell (Abbildung 5.4) für jedes Portal separat dargestellt.

Im Folgenden werden die auf Basis der $m = 5$ imputierten Tabellen und anhand der in Anhang A beschriebenen Poolingregeln (Gleichungen A.4–A.9) ermittelten Ergebnisse dargestellt. Diese deuten auf gute Konstruktvalidität (Abschnitte C.1, C.2 und C.3) sowie strukturelle Validität (Abschnitte C.4 und C.5) des Pfadmodells hin. Es sind jedoch nicht alle angenommenen Abhängigkeiten zwischen den Konstrukten (bzw. latenten Variablen) signifikant (siehe Abschnitt C.4).

Die Reliabilität der Konstrukte ist mit Ausnahme von VERT (Vertrauen) recht hoch (Abschnitt C.6). Eine Reformulierung der Fragen zu VERT wäre bei zukünftigen Untersuchungen zu erwägen.

Für Portal P3 fallen die Validierungsergebnisse zum Teil etwas schlechter aus. Die Stichprobengröße für P3 ist im Vergleich zu den anderen Portalen deutlich geringer (siehe Spalte „abgeschlossen“ in Tabelle 6.11), was möglicherweise Ursache für die etwas schlechteren Werte sein könnte.

C.1 Signifikanz der Faktorladungen

Die t-Statistiken (Tabellen C.1, C.2, C.3, C.4) zu den Faktorladungen weisen bei fast allen Portalen signifikante Abhängigkeiten (zweiseitiger Test; $p \leq 0,05$) für alle Faktoren auf und deuten damit auf eine gute Konvergenzvalidität. Ausnahme ist das Konstrukt RISK (Risiko) bei den Portalen P2 und P3. Bei P2 ist die Faktorladung des Items RISK2 nicht signifikant (Tabelle C.2) und bei P3 sind die Faktorladungen der Items RISK1 und RISK2 nicht signifikant (Tabelle C.3).

Der relative Anstieg in den Varianzen aufgrund der imputierten Werte (Indikator r) ist bei allen Portalen für alle Faktorladungen gering. Die relative Effizienz der Imputation (Indikator RE) ist für $m = 5$ im Vergleich zu $m \rightarrow \infty$ bei allen Portalen für alle Faktorladungen größer als 0,97, beträgt also mehr als 97% (zur Bedeutung der Kennzahlen siehe Anhang A).

Tabelle C.1: t-Statistik zu den gepoolten Faktorladungen (\bar{Q}) für P1

Faktor \leftarrow Konstrukt	\bar{Q}	\bar{U}	\hat{B}	\hat{T}	t	[ν]	p	r	$\hat{\lambda}$	RE
INFQ1 \leftarrow INFQ	0,3365	0,0007	<0,0001	0,0007	12,6752	4182	<0,01	0,0319	0,0005	0,9999
INFQ2 \leftarrow INFQ	0,291	0,0005	<0,0001	0,0005	13,4329	4695	<0,01	0,0301	0,0004	0,9999
INFQ3 \leftarrow INFQ	0,2839	0,0006	0,0001	0,0008	10,1795	142	<0,01	0,2011	0,0126	0,9975
INFQ4 \leftarrow INFQ	0,2966	0,0004	0,0001	0,0005	13,4483	90	<0,01	0,2657	0,0191	0,9962
LOYA1 \leftarrow LOYA	0,5036	0,0001	0,0001	0,0002	34,8975	29	<0,01	0,581	0,0501	0,9901
LOYA2 \leftarrow LOYA	0,5359	0,0002	<0,0001	0,0002	35,3624	373	<0,01	0,1154	0,005	0,999
RISK1 \leftarrow RISK	0,6706	0,0164	0,0146	0,0339	3,6445	15	<0,01	1,0663	0,0823	0,9838
RISK2 \leftarrow RISK	0,4415	0,0219	0,0187	0,0444	2,0961	15	0,05	1,022	0,0801	0,9842
SYSQ1 \leftarrow PREQ	0,4135	0,0005	<0,0001	0,0006	17,4636	1903	<0,01	0,048	0,001	0,9998
SYSQ2 \leftarrow PREQ	0,3573	0,001	<0,0001	0,0011	10,9936	2200	<0,01	0,0445	0,0009	0,9998
SYSQ3 \leftarrow PREQ	0,3822	0,0008	<0,0001	0,0008	13,2236	1953	<0,01	0,0474	0,001	0,9998
SYSQ4 \leftarrow TECQ	0,3777	0,002	0,0001	0,0021	8,236	1117	<0,01	0,0636	0,0017	0,9997
SYSQ5 \leftarrow TECQ	0,4217	0,0017	0,0003	0,0021	9,2709	103	<0,01	0,2448	0,0169	0,9966
SYSQ6 \leftarrow TECQ	0,4121	0,0016	0,0003	0,0019	9,5463	148	<0,01	0,1961	0,0121	0,9976
VERT1 \leftarrow VERT	0,6881	0,003	0,001	0,0042	10,608	51	<0,01	0,3872	0,0317	0,9937
VERT2 \leftarrow VERT	0,477	0,0019	0,0004	0,0024	9,8088	101	<0,01	0,2471	0,0172	0,9966
WISS1 \leftarrow WISS	0,5774	0,0005	0,0001	0,0006	23,7196	106	<0,01	0,24	0,0165	0,9967
WISS2 \leftarrow WISS	0,4932	0,0003	<0,0001	0,0003	27,2502	318	<0,01	0,1261	0,0059	0,9988

Tabelle C.2: t-Statistik zu den gepoolten Faktorladungen (\bar{Q}) für P2

Faktor \leftarrow Konstrukt	\bar{Q}	\bar{U}	\hat{B}	\hat{T}	t	[v]	p	r	$\hat{\lambda}$	RE
INFQ1 \leftarrow INFQ	0,3197	0,0007	0,0001	0,0008	11,5461	638	<0,01	0,086	0,003	0, 9994
INFQ2 \leftarrow INFQ	0,3093	0,0004	<0,0001	0,0005	14,0525	738	<0,01	0,0795	0,0026	0, 9995
INFQ3 \leftarrow INFQ	0,2396	0,0008	0,0002	0,001	7,5312	117	<0,01	0,2262	0,0151	0, 997
INFQ4 \leftarrow INFQ	0,3558	0,0011	0,0001	0,0013	10,0539	276	<0,01	0,1368	0,0067	0, 9987
LOYA1 \leftarrow LOYA	0,5068	0,0007	<0,0001	0,0008	18,2916	6739	<0,01	0,025	0,0003	0, 9999
LOYA2 \leftarrow LOYA	0,5951	0,0011	<0,0001	0,0011	17,642	3813	<0,01	0,0335	0,0005	0, 9999
RISK1 \leftarrow RISK	0,6211	0,071	0,0215	0,0968	1,9966	56	0,05	0,3625	0,0291	0,9942
RISK2 \leftarrow RISK	0,5227	0,0675	0,0367	0,1116	1,5648	25	0,13	0,6526	0,056	0,9889
SYSQ1 \leftarrow PREQ	0,3894	0,0004	<0,0001	0,0004	19,5063	25651	<0,01	0,0126	0,0001	1
SYSQ2 \leftarrow PREQ	0,3949	0,0005	<0,0001	0,0005	18,2527	16677	<0,01	0,0157	0,0001	1
SYSQ3 \leftarrow PREQ	0,3394	0,0005	<0,0001	0,0006	14,0953	1064	<0,01	0,0653	0,0018	0, 9996
SYSQ4 \leftarrow TECQ	0,4381	0,0009	0,0001	0,001	13,8222	420	<0,01	0,1081	0,0045	0, 9991
SYSQ5 \leftarrow TECQ	0,3853	0,0008	<0,0001	0,0008	13,5633	1141	<0,01	0,0629	0,0017	0, 9997
SYSQ6 \leftarrow TECQ	0,3722	0,0017	0,0002	0,0019	8,5446	370	<0,01	0,1159	0,0051	0, 999
VERT1 \leftarrow VERT	0,6874	0,0031	0,001	0,0044	10,3827	48	<0,01	0,4023	0,0332	0, 9934
VERT2 \leftarrow VERT	0,5022	0,002	0,0011	0,0033	8,7146	25	<0,01	0,6585	0,0565	0, 9888
WISS1 \leftarrow WISS	0,6022	0,0007	0,0001	0,0008	21,3051	415	<0,01	0,1088	0,0045	0, 9991
WISS2 \leftarrow WISS	0,4944	0,0005	0,0001	0,0005	21,2307	304	<0,01	0,1295	0,0061	0, 9988

Tabelle C.3: t-Statistik zu den gepoolten Faktorladungen (\bar{Q}) für P3

Faktor \leftarrow Konstrukt	\bar{Q}	\bar{U}	\hat{B}	\hat{T}	t	[ν]	p	r	$\hat{\lambda}$	RE
INFQ1 \leftarrow INFQ	0,3066	0,0006	0,0001	0,0007	11,2313	88	<0,01	0,2706	0,0196	0, 9961
INFQ2 \leftarrow INFQ	0,3206	0,0008	0,0001	0,0008	11,0529	625	<0,01	0,0869	0,0031	0, 9994
INFQ3 \leftarrow INFQ	0,2825	0,0013	0,0001	0,0014	7,5463	385	<0,01	0,1134	0,0049	0, 999
INFQ4 \leftarrow INFQ	0,2841	0,0008	0,0001	0,001	9,2077	215	<0,01	0,1578	0,0085	0, 9983
LOYA1 \leftarrow LOYA	0,5443	0,0006	0,0001	0,0007	20,8931	199	<0,01	0,1652	0,0092	0, 9982
LOYA2 \leftarrow LOYA	0,5297	0,0004	0,0001	0,0006	22,4859	83	<0,01	0,2804	0,0206	0, 9959
RISK1 \leftarrow RISK	0,291	0,1753	0,3028	0,5387	0,3965	8	0,70	2,0729	0,1124	0, 978
RISK2 \leftarrow RISK	0,6777	0,1714	0,1239	0,32	1,198	18	0,25	0,8675	0,0713	0, 9859
SYSQ1 \leftarrow PREQ	0,362	0,0005	0,0001	0,0006	15,3296	177	<0,01	0,1764	0,0102	0, 998
SYSQ2 \leftarrow PREQ	0,3786	0,0006	0,0001	0,0008	13,5209	130	<0,01	0,2123	0,0137	0, 9973
SYSQ3 \leftarrow PREQ	0,3778	0,0006	0,0001	0,0007	14,2869	297	<0,01	0,1311	0,0063	0, 9987
SYSQ4 \leftarrow TECQ	0,4203	0,0011	0,0002	0,0013	11,4516	131	<0,01	0,2111	0,0136	0, 9973
SYSQ5 \leftarrow TECQ	0,392	0,0012	<0,0001	0,0012	11,1733	2938	<0,01	0,0383	0,0007	0, 9999
SYSQ6 \leftarrow TECQ	0,3415	0,0017	0,0001	0,0019	7,8756	787	<0,01	0,0767	0,0024	0, 9995
VERT1 \leftarrow VERT	0,7276	0,0056	0,0043	0,0107	7,0202	17	<0,01	0,9237	0,0747	0, 9853
VERT2 \leftarrow VERT	0,451	0,0054	0,0025	0,0084	4,9247	31	<0,01	0,5578	0,048	0, 9905
WISS1 \leftarrow WISS	0,7642	0,0038	0,0001	0,0039	12,2947	5330	<0,01	0,0282	0,0004	0, 9999
WISS2 \leftarrow WISS	0,362	0,003	0,0001	0,0031	6,4871	2645	<0,01	0,0405	0,0007	0,9999

Tabelle C.4: t-Statistik zu den gepoolten Faktorladungen (\bar{Q}) für P4

Faktor \leftarrow Konstrukt	\bar{Q}	\bar{U}	\hat{B}	\hat{T}	t	[ν]	p	r	$\hat{\lambda}$	RE
INFQ1 \leftarrow INFQ	0,3432	0,0005	0,0001	0,0005	14,8256	168	<0,01	0,1819	0,0107	0, 9979
INFQ2 \leftarrow INFQ	0,2938	0,0004	0,0001	0,0005	12,9044	183	<0,01	0,1731	0,0099	0, 998
INFQ3 \leftarrow INFQ	0,2401	0,0005	0,0001	0,0006	9,6007	52	<0,01	0,3814	0,0311	0, 9938
INFQ4 \leftarrow INFQ	0,343	0,0007	0,0001	0,0007	12,6224	313	<0,01	0,1274	0,006	0, 9988
LOYA1 \leftarrow LOYA	0,5017	0,0002	<0,0001	0,0002	32,5725	879	<0,01	0,0723	0,0022	0,9996
LOYA2 \leftarrow LOYA	0,5467	0,0004	<0,0001	0,0004	26,7822	3320	<0,01	0,036	0,0006	0,9999
RISK1 \leftarrow RISK	0,5951	0,0047	0,0012	0,0061	7,5895	71	<0,01	0,3107	0,0238	0,9953
RISK2 \leftarrow RISK	0,5522	0,0045	0,0034	0,0085	5,9914	17	<0,01	0,9065	0,0737	0,9855
SYSQ1 \leftarrow PREQ	0,4098	0,0004	<0,0001	0,0005	19,1599	2444	<0,01	0,0422	0,0008	0,9998
SYSQ2 \leftarrow PREQ	0,3692	0,0006	<0,0001	0,0007	14,0921	1017	<0,01	0,0669	0,0019	0,9996
SYSQ3 \leftarrow PREQ	0,3798	0,0006	<0,0001	0,0006	14,9932	754	<0,01	0,0785	0,0025	0,9995
SYSQ4 \leftarrow TECQ	0,3795	0,0017	0,0007	0,0025	7,5647	36	<0,01	0,4911	0,0419	0,9917
SYSQ5 \leftarrow TECQ	0,43	0,0016	0,0005	0,0021	9,2793	53	<0,01	0,3787	0,0308	0,9939
SYSQ6 \leftarrow TECQ	0,4441	0,0019	0,0002	0,0021	9,6176	476	<0,01	0,1009	0,004	0, 9992
VERT1 \leftarrow VERT	0,7185	0,0025	0,003	0,0061	9,1655	11	<0,01	1,4542	0,0978	0, 9808
VERT2 \leftarrow VERT	0,4693	0,0018	0,0022	0,0044	7,0766	11	<0,01	1,4483	0,0976	0, 9809
WISS1 \leftarrow WISS	0,5617	0,0004	<0,0001	0,0004	27,7353	238	<0,01	0,1488	0,0077	0,9985
WISS2 \leftarrow WISS	0,5147	0,0003	<0,0001	0,0003	28,7376	932	<0,01	0,0701	0,0021	0,9996

C.2 Faktorladungen (*Crossloadings*)

Die gepoolten Faktorladungsmatrizen (Tabellen C.5, C.7, C.9, C.11) zeigen bei allen Portalen, dass die Ladungen der Messvariablen auf ihre zugeordneten Konstrukte deutlich höher sind, als die Ladungen der nicht zugeordneten Messvariablen. Eine leichte Abweichung davon ist bei Portal P3 zu sehen, bei welchem die Unterschiede zwischen den Ladungen der Messvariablen zu PREQ (Präsentationsqualität: Items SYSQ1–SYSQ3) und der Messvariablen zu TECQ (technische Qualität: Items SYSQ4–SYSQ6) nicht so groß sind.

Die zugehörigen Standardabweichungen, die sich aus den $m = 5$ Punktschätzern für die Ladungen ergeben, sind in den Tabellen C.6, C.8, C.10 und C.12 enthalten.

Tabelle C.5: Gepoolte Faktorladungsmatrix zu P1

	INFQ	LOYA	PREQ	RISK	TECQ	VERT	WISS
INFQ1	0,8291	0,5685	0,5917	0,2135	0,5914	0,5432	0,6596
INFQ2	0,8375	0,5588	0,6632	0,2121	0,6092	0,5703	0,4572
INFQ3	0,8342	0,4911	0,5546	0,192	0,5595	0,466	0,5441
INFQ4	0,8106	0,5354	0,5684	0,241	0,6398	0,5703	0,4659
LOYA1	0,5859	0,9596	0,4425	0,2425	0,4805	0,633	0,6459
LOYA2	0,6668	0,9643	0,5321	0,2518	0,5742	0,698	0,6593
RISK1	0,2599	0,2476	0,2507	0,9268	0,151	0,2603	0,2238
RISK2	0,1817	0,1868	0,2145	0,8305	0,1561	0,1781	0,143
SYSQ1	0,6369	0,4878	0,9275	0,2292	0,5818	0,4631	0,6213
SYSQ2	0,6731	0,3821	0,8128	0,2242	0,6035	0,4833	0,4517
SYSQ3	0,5609	0,4441	0,8532	0,2469	0,551	0,4734	0,5189
SYSQ4	0,6096	0,4541	0,5391	0,1852	0,798	0,4417	0,4101
SYSQ5	0,6144	0,486	0,5528	0,085	0,8568	0,5079	0,4794
SYSQ6	0,5704	0,4192	0,5575	0,1662	0,8178	0,4467	0,4932
VERT1	0,6555	0,6779	0,5336	0,2393	0,5396	0,9056	0,6189
VERT2	0,4178	0,4725	0,3729	0,21	0,4067	0,7889	0,3949
WISS1	0,6573	0,6617	0,6445	0,205	0,5749	0,606	0,9441
WISS2	0,5441	0,6011	0,4959	0,2133	0,4642	0,537	0,9224

Tabelle C.6: Standardabweichungen ($\sqrt{\hat{B}}$) zu Tabelle C.5 für P1

	INFQ	LOYA	PREQ	RISK	TECQ	VERT	WISS
INFQ1	0,007	0,0171	0,0047	0,0644	0,0131	0,0271	0,0111
INFQ2	0,0066	0,0207	0,0029	0,0652	0,0117	0,0255	0,0212
INFQ3	0,0087	0,0223	0,0193	0,0545	0,0378	0,0119	0,0271
INFQ4	0,0065	0,0159	0,0126	0,0716	0,0055	0,0175	0,0146
LOYA1	0,0144	0,0076	0,0064	0,0645	0,0179	0,0293	0,0105
LOYA2	0,0209	0,0076	0,0144	0,0966	0,022	0,0252	0,0152
RISK1	0,0628	0,0457	0,0427	0,0533	0,0393	0,076	0,0434
RISK2	0,0946	0,1164	0,0947	0,0901	0,0723	0,1399	0,1222
SYSQ1	0,0094	0,0076	0,0017	0,0567	0,0028	0,0074	0,0177
SYSQ2	0,0045	0,0127	0,0044	0,0597	0,0078	0,0197	0,019
SYSQ3	0,0114	0,0191	0,0032	0,0562	0,0096	0,0277	0,0098
SYSQ4	0,011	0,0202	0,0121	0,0319	0,0069	0,0142	0,0233
SYSQ5	0,0091	0,0166	0,0082	0,0526	0,0052	0,0215	0,0161
SYSQ6	0,0338	0,0137	0,0174	0,0408	0,0218	0,0336	0,0288
VERT1	0,0161	0,0061	0,0083	0,102	0,0108	0,0067	0,014
VERT2	0,0347	0,0658	0,0375	0,0728	0,0435	0,0356	0,0507
WISS1	0,0144	0,0062	0,0119	0,0844	0,0133	0,023	0,0021
WISS2	0,0309	0,0179	0,0231	0,0462	0,0184	0,0304	0,007

Tabelle C.7: Gepoolte Faktorladungsmatrix zu P2

	INFQ	LOYA	PREQ	RISK	TECQ	VERT	WISS
INFQ1	0,8471	0,4384	0,6548	0,0644	0,5789	0,5239	0,5047
INFQ2	0,8453	0,4218	0,6995	0,109	0,5791	0,5206	0,4651
INFQ3	0,7364	0,3145	0,4544	0,1153	0,3663	0,3946	0,3646
INFQ4	0,8177	0,5203	0,528	0,1323	0,461	0,5111	0,6175
LOYA1	0,4325	0,8906	0,4264	0,1108	0,3644	0,5274	0,5811
LOYA2	0,5248	0,9219	0,527	0,1474	0,4493	0,6519	0,6503
RISK1	0,1126	0,1226	0,0238	0,8688	0,074	0,1574	0,0911
RISK2	0,0869	0,1132	0,0274	0,8406	0,0721	0,0411	0,0632
SYSQ1	0,6963	0,4929	0,9033	0,0446	0,5994	0,5792	0,5031
SYSQ2	0,6632	0,4884	0,9067	0,0204	0,6216	0,589	0,5087
SYSQ3	0,5579	0,4267	0,8548	0,0346	0,5911	0,5026	0,4443
SYSQ4	0,5581	0,4145	0,5866	0,0953	0,8771	0,5595	0,4455
SYSQ5	0,5572	0,4068	0,6345	0,0246	0,8664	0,5199	0,3669
SYSQ6	0,4207	0,3049	0,4779	0,1197	0,7572	0,4387	0,4163
VERT1	0,5788	0,6374	0,5648	0,1398	0,5654	0,8872	0,6261
VERT2	0,4109	0,4308	0,4761	0,0676	0,4354	0,7751	0,3703
WISS1	0,6319	0,6543	0,578	0,1023	0,4821	0,5913	0,9284
WISS2	0,4692	0,5811	0,4033	0,069	0,4075	0,5285	0,8916

Tabelle C.8: Standardabweichungen ($\sqrt{\hat{B}}$) zu Tabelle C.7 für P2

	INFQ	LOYA	PREQ	RISK	TECQ	VERT	WISS
INFQ1	0,0058	0,0516	0,0073	0,0509	0,0155	0,0309	0,0214
INFQ2	0,002	0,0499	0,0061	0,0396	0,0152	0,027	0,0207
INFQ3	0,0217	0,0132	0,0223	0,0374	0,0209	0,0231	0,0208
INFQ4	0,0112	0,0193	0,014	0,0997	0,0165	0,0327	0,0087
LOYA1	0,0361	0,0037	0,0311	0,0542	0,0396	0,0281	0,0112
LOYA2	0,0357	0,0007	0,0275	0,0905	0,0333	0,0425	0,0217
RISK1	0,0915	0,0779	0,1022	0,0965	0,0872	0,1026	0,0946
RISK2	0,035	0,0781	0,062	0,0713	0,0399	0,0742	0,0736
SYSQ1	0,0082	0,0286	0,004	0,0916	0,0153	0,0265	0,0102
SYSQ2	0,0083	0,0332	0,0027	0,0985	0,0133	0,0367	0,02
SYSQ3	0,0176	0,0215	0,0021	0,0662	0,0104	0,028	0,0107
SYSQ4	0,013	0,0467	0,009	0,0507	0,0057	0,0455	0,0319
SYSQ5	0,0087	0,0485	0,0058	0,0574	0,0035	0,0393	0,0264
SYSQ6	0,022	0,0305	0,0195	0,0675	0,0082	0,0394	0,0375
VERT1	0,0091	0,017	0,0165	0,0657	0,0277	0,0157	0,0173
VERT2	0,0543	0,0731	0,062	0,1245	0,0651	0,0292	0,043
WISS1	0,014	0,0206	0,0143	0,1189	0,0256	0,011	0,0023
WISS2	0,0101	0,0167	0,023	0,0638	0,0349	0,0109	0,0049

Tabelle C.9: Gepoolte Faktorladungsmatrix zu P3

	INFQ	LOYA	PREQ	RISK	TECQ	VERT	WISS
INFQ1	0,8881	0,5081	0,5538	0,0414	0,5132	0,5006	0,6131
INFQ2	0,8475	0,51	0,6939	0,0625	0,6516	0,5711	0,594
INFQ3	0,7832	0,3565	0,4557	0,0963	0,4452	0,4682	0,5624
INFQ4	0,8254	0,5239	0,6381	0,0584	0,6	0,4314	0,5902
LOYA1	0,4923	0,933	0,5864	0,1003	0,5589	0,6687	0,6744
LOYA2	0,5691	0,9291	0,6157	0,037	0,6268	0,6797	0,6183
RISK1	0,1197	0,0933	0,1243	0,5549	0,0746	0,1155	0,1367
RISK2	0,0224	0,0588	0,0114	0,8775	0,0512	0,0348	0,1234
SYSQ1	0,6545	0,5995	0,9136	0,0165	0,7082	0,5253	0,5787
SYSQ2	0,6366	0,5553	0,8935	0,0205	0,7624	0,5836	0,5768
SYSQ3	0,5942	0,5771	0,8762	0,0604	0,6995	0,5279	0,6205
SYSQ4	0,56	0,6016	0,6851	0,1063	0,9007	0,6059	0,5902
SYSQ5	0,5985	0,5704	0,7927	0,074	0,8923	0,553	0,5536
SYSQ6	0,5653	0,4678	0,616	0,0477	0,7951	0,4668	0,5119
VERT1	0,6102	0,7113	0,6263	0,0735	0,6156	0,9112	0,6247
VERT2	0,3235	0,4487	0,3383	0,0629	0,4058	0,7382	0,3137
WISS1	0,7349	0,7417	0,7216	0,1438	0,7146	0,6511	0,9506
WISS2	0,3961	0,3525	0,306	0,0458	0,2592	0,2799	0,7554

Tabelle C.10: Standardabweichungen (\sqrt{B}) zu Tabelle C.9 für P3

	INFQ	LOYA	PREQ	RISK	TECQ	VERT	WISS
INFQ1	0,011	0,0343	0,0205	0,0698	0,0182	0,031	0,0275
INFQ2	0,0105	0,0371	0,0129	0,1327	0,0023	0,0164	0,0124
INFQ3	0,0183	0,0298	0,0221	0,1198	0,0238	0,044	0,0287
INFQ4	0,0126	0,0345	0,0202	0,0945	0,014	0,0336	0,0153
LOYA1	0,0209	0,0054	0,0135	0,0915	0,0225	0,0243	0,0167
LOYA2	0,0321	0,0047	0,0215	0,0832	0,0409	0,03	0,013
RISK1	0,0674	0,0771	0,0937	0,5151	0,114	0,0987	0,0731
RISK2	0,027	0,0693	0,1581	0,0833	0,0534	0,0592	0,0967
SYSQ1	0,0107	0,0175	0,0033	0,1809	0,0116	0,013	0,0069
SYSQ2	0,012	0,0206	0,002	0,1762	0,0156	0,0348	0,0142
SYSQ3	0,0103	0,0092	0,0055	0,1899	0,0109	0,0176	0,0058
SYSQ4	0,0253	0,022	0,01	0,0326	0,008	0,0395	0,0162
SYSQ5	0,0081	0,033	0,0128	0,1608	0,0077	0,0077	0,0106
SYSQ6	0,0192	0,027	0,0155	0,0253	0,0149	0,0296	0,0104
VERT1	0,0163	0,0079	0,0061	0,0344	0,0105	0,0146	0,0152
VERT2	0,0598	0,1055	0,0752	0,1654	0,0379	0,0837	0,0764
WISS1	0,0167	0,0186	0,0113	0,1575	0,006	0,0192	0,0031
WISS2	0,0144	0,0172	0,0132	0,0222	0,021	0,0221	0,0121

Tabelle C.11: Gepoolte Faktorladungsmatrix zu P4

	INFQ	LOYA	PREQ	RISK	TECQ	VERT	WISS
INFQ1	0,8671	0,46	0,5769	0,2815	0,4848	0,6036	0,5715
INFQ2	0,7848	0,4959	0,6341	0,2683	0,5516	0,5227	0,4727
INFQ3	0,7726	0,3208	0,3834	0,0837	0,2755	0,3899	0,4782
INFQ4	0,8345	0,499	0,5394	0,2311	0,5228	0,5663	0,6277
LOYA1	0,5261	0,9496	0,5092	0,2184	0,3994	0,5747	0,5733
LOYA2	0,5272	0,9577	0,5562	0,3188	0,4643	0,6382	0,6076
RISK1	0,2717	0,2598	0,2955	0,8798	0,3468	0,4258	0,1836
RISK2	0,2058	0,2343	0,293	0,8636	0,3177	0,3844	0,1946
SYSQ1	0,5656	0,5614	0,8999	0,3146	0,5589	0,5672	0,519
SYSQ2	0,5686	0,3918	0,831	0,2963	0,6311	0,5611	0,4006
SYSQ3	0,5783	0,4864	0,8541	0,2589	0,5366	0,5302	0,4959
SYSQ4	0,4325	0,2931	0,4486	0,2403	0,7915	0,436	0,3274
SYSQ5	0,5079	0,3583	0,5562	0,3605	0,8406	0,4497	0,3562
SYSQ6	0,4308	0,4238	0,5775	0,3047	0,7601	0,4861	0,3974
VERT1	0,6748	0,6302	0,6651	0,3806	0,5428	0,9015	0,6828
VERT2	0,3488	0,3856	0,3446	0,401	0,3924	0,7449	0,256
WISS1	0,6417	0,6075	0,5127	0,2079	0,4452	0,5966	0,9353
WISS2	0,5925	0,5411	0,5071	0,1985	0,398	0,5392	0,9223

Tabelle C.12: Standardabweichungen ($\sqrt{\hat{B}}$) zu Tabelle C.11 für P4

	INFQ	LOYA	PREQ	RISK	TECQ	VERT	WISS
INFQ1	0,006	0,0197	0,0074	0,0376	0,0168	0,0207	0,0177
INFQ2	0,0054	0,0164	0,0041	0,0198	0,0047	0,0244	0,027
INFQ3	0,0111	0,0363	0,0224	0,0512	0,0204	0,0451	0,0329
INFQ4	0,0047	0,0174	0,011	0,0445	0,0115	0,0401	0,0205
LOYA1	0,0129	0,003	0,0183	0,0618	0,012	0,0254	0,0112
LOYA2	0,0172	0,0026	0,0212	0,0791	0,0165	0,0264	0,0162
RISK1	0,0422	0,0792	0,0288	0,0514	0,0596	0,0409	0,0361
RISK2	0,0385	0,0757	0,0385	0,0372	0,0511	0,0591	0,0861
SYSQ1	0,007	0,0134	0,0015	0,0099	0,003	0,0148	0,0225
SYSQ2	0,0077	0,0229	0,0021	0,0205	0,0057	0,0247	0,0233
SYSQ3	0,0133	0,0206	0,004	0,0251	0,0119	0,0162	0,0162
SYSQ4	0,0049	0,0273	0,0134	0,0475	0,0127	0,0242	0,0313
SYSQ5	0,0158	0,0117	0,0106	0,0525	0,0099	0,0151	0,0317
SYSQ6	0,0127	0,0169	0,0111	0,0463	0,0118	0,0159	0,0285
VERT1	0,0071	0,0059	0,0131	0,0393	0,0146	0,0173	0,0088
VERT2	0,0353	0,0909	0,0437	0,068	0,0384	0,062	0,0554
WISS1	0,0207	0,0224	0,0148	0,0503	0,011	0,0165	0,0028
WISS2	0,0248	0,0192	0,0281	0,045	0,0297	0,0074	0,0052

C.3 AVE- und Korrelationsmatrizen

Die gepoolten AVE-Werte (\bar{Q}) und die zugehörigen Standardabweichungen ($\sqrt{\hat{B}}$) sind in den Tabellen C.13, C.16, C.19 und C.22 aufgeführt. Die gepoolten AVE- und Korrelationsmatrizen sind in den Tabellen C.14, C.17, C.20 und C.23 dargestellt. Die Standardabweichungen, die sich aus den $m = 5$ Punktschätzern für die Korrelationskoeffizienten ergeben, sind in den Tabellen C.15, C.18, C.21 und C.24 dargestellt.

Für alle Portale ist \sqrt{AVE} jedes Konstrukts größer, als dessen Korrelationen mit jedem anderen Konstrukt. Dies deutet auf eine gute Diskriminanzvalidität des Messmodells hin. Eine leichte Abweichung davon ist bei Portal P3 zu sehen, bei welchem die Unterschiede zwischen den Korrelationskoeffizienten für PREQ (Präsentationsqualität) und TECQ (technische Qualität) nicht so groß sind.

Tabelle C.13: Gepoolte AVE-Werte (\bar{Q}) für P1

Konstrukt	\bar{Q}	$\sqrt{\hat{B}}$
INFQ	0,6855	0,0066
LOYA	0,9254	0,0145
PREQ	0,7496	0,0018
RISK	0,7788	0,0622
TECQ	0,6801	0,0122
VERT	0,7218	0,0267
WISS	0,8711	0,0082

Tabelle C.14: Gepoolte AVE- und Korrelationsmatrix für P1

$\sqrt{\bar{Q}_{\text{AVE}}}$ und $\bar{Q}_{\text{Korrelation}}$	INFQ	LOYA	PREQ	RISK	TECQ	VERT	WISS
INFQ	0,8279						
LOYA	0,6524	0,9620					
PREQ	0,7183	0,508	0,8658				
RISK	0,2597	0,2568	0,2692	0,8825			
TECQ	0,7246	0,5496	0,6667	0,1745	0,8247		
VERT	0,6505	0,6927	0,5452	0,2636	0,5657	0,8496	
WISS	0,6478	0,6785	0,6167	0,2237	0,5609	0,6146	0,9333

Die $\sqrt{\text{AVE}}$ -Werte sind in Fettdruck auf der Diagonalen abgebildet.

Tabelle C.15: Standardabweichungen ($\sqrt{\hat{B}}$) zu Tabelle C.14 für P1

	INFQ	LOYA	PREQ	RISK	TECQ	VERT	WISS
INFQ	0						
LOYA	0,0126	0					
PREQ	0,0069	0,0111	0				
RISK	0,0697	0,08	0,0622	0			
TECQ	0,0087	0,0168	0,0051	0,0497	0		
VERT	0,0165	0,0238	0,0159	0,0961	0,0113	0	
WISS	0,02	0,0086	0,0147	0,0695	0,0143	0,0226	0

Tabelle C.16: Gepoolte AVE-Werte (\bar{Q}) für P2

Konstrukt	\bar{Q}	$\sqrt{\hat{B}}$
INFQ	0,6609	0,0057
LOYA	0,8216	0,0032
PREQ	0,7896	0,0044
RISK	0,7364	0,045
TECQ	0,6978	0,0033
VERT	0,6944	0,0187
WISS	0,8285	0,0047

Tabelle C.17: Gepoolte AVE- und Korrelationsmatrix für P2

$\sqrt{Q_{AVE}}$ und $\bar{Q}_{Korrelation}$	INFQ	LOYA	PREQ	RISK	TECQ	VERT	WISS
INFQ	0,813						
LOYA	0,5315	0,9064					
PREQ	0,7224	0,5297	0,8886				
RISK	0,1296	0,1438	0,0371	0,8581			
TECQ	0,6159	0,4521	0,6795	0,0956	0,8353		
VERT	0,6054	0,6553	0,6288	0,1305	0,6089	0,8333	
WISS	0,6126	0,6815	0,5476	0,0955	0,4919	0,6174	0,9102

Die \sqrt{AVE} -Werte sind in Fettdruck auf der Diagonalen abgebildet.

Tabelle C.18: Standardabweichungen ($\sqrt{\hat{B}}$) zu Tabelle C.17 für P2

	INFQ	LOYA	PREQ	RISK	TECQ	VERT	WISS
INFQ	0						
LOYA	0,0394	0					
PREQ	0,0099	0,0316	0				
RISK	0,0611	0,0794	0,0946	0			
TECQ	0,0155	0,0386	0,0115	0,0621	0		
VERT	0,0251	0,0372	0,0329	0,1074	0,0452	0	
WISS	0,009	0,0173	0,0127	0,0976	0,0293	0,0047	0

Tabelle C.19: Gepoolte AVE-Werte (\bar{Q}) für P3

Konstrukt	\bar{Q}	$\sqrt{\hat{B}}$
INFQ	0,7005	0,0117
LOYA	0,8668	0,0081
PREQ	0,8002	0,005
RISK	0,6478	0,1569
TECQ	0,7466	0,0089
VERT	0,6905	0,0466
WISS	0,7372	0,0091

Tabelle C.20: Gepoolte AVE- und Korrelationsmatrix für P3

$\sqrt{\bar{Q}_{\text{AVE}}}$ und $\bar{Q}_{\text{Korrelation}}$	INFQ	LOYA	PREQ	RISK	TECQ	VERT	WISS
INFQ	0,837						
LOYA	0,5695	0,931					
PREQ	0,7024	0,6454	0,8946				
RISK	0,078	0,0745	0,0368	0,8049			
TECQ	0,6629	0,6363	0,8092	0,09	0,8641		
VERT	0,5922	0,7242	0,6109	0,0847	0,6315	0,831	
WISS	0,6597	0,6974	0,6127	0,1108	0,5823	0,6116	0,8586

Die $\sqrt{\text{AVE}}$ -Werte sind in Fettdruck auf der Diagonalen abgebildet.

Tabelle C.21: Standardabweichungen ($\sqrt{\hat{B}}$) zu Tabelle C.20 für P3

	INFQ	LOYA	PREQ	RISK	TECQ	VERT	WISS
INFQ	0						
LOYA	0,0258	0					
PREQ	0,0076	0,0186	0				
RISK	0,0565	0,0935	0,2015	0			
TECQ	0,0082	0,032	0,0054	0,0745	0		
VERT	0,0145	0,0177	0,0176	0,0965	0,016	0	
WISS	0,051	0,0098	0,0688	0,1015	0,0816	0,0074	0

Tabelle C.22: Gepoolte AVE-Werte (\bar{Q}) für P4

Konstrukt	\bar{Q}	$\sqrt{\hat{B}}$
INFQ	0,6654	0,0053
LOYA	0,9094	0,0052
PREQ	0,7433	0,0016
RISK	0,7615	0,0715
TECQ	0,6371	0,0054
VERT	0,6854	0,0346
WISS	0,8627	0,0071

Tabelle C.23: Gepoolte AVE- und Korrelationsmatrix für P4

\sqrt{Q}_{AVE} und $\bar{Q}_{Korrelation}$	INFQ	LOYA	PREQ	RISK	TECQ	VERT	WISS
INFQ	0,8157						
LOYA	0,5522	0,9536					
PREQ	0,6614	0,5595	0,8621				
RISK	0,2748	0,2838	0,3367	0,8727			
TECQ	0,574	0,4542	0,6658	0,3829	0,7982		
VERT	0,6493	0,6372	0,6412	0,4649	0,5755	0,8279	
WISS	0,6654	0,6198	0,549	0,2191	0,4549	0,6126	0,9288

Die \sqrt{AVE} -Werte sind in Fettdruck auf der Diagonalen abgebildet.

Tabelle C.24: Standardabweichungen ($\sqrt{\hat{B}}$) zu Tabelle C.23 für P4

	INFQ	LOYA	PREQ	RISK	TECQ	VERT	WISS
INFQ	0						
LOYA	0,015	0					
PREQ	0,0057	0,0198	0				
RISK	0,0324	0,0719	0,0192	0			
TECQ	0,0103	0,0134	0,0037	0,0387	0		
VERT	0,022	0,0255	0,0168	0,0352	0,007	0	
WISS	0,0259	0,0127	0,0232	0,0409	0,0209	0,0116	0

C.4 Abhangigkeiten zwischen den Konstrukten

Die t-Statistiken (Tabellen C.25, C.26, C.27 und C.28) zu den Pfadkoeffizienten (Strukturmodell) weisen nicht fur alle Pfade signifikante Werte (zweiseitiger Test; $p \leq 0,05$) bei den Portalen auf. Ermittelte signifikante Abhangigkeiten zwischen den Konstrukten (bzw. latenten Variablen) sind auch nicht bei allen Portalen identisch. Strukturelle Validitat des Pfadmodells ist damit nicht fur alle angenommenen Beziehungen zwischen den latenten Variablen (siehe Abbildung 5.4) gegeben.

Der relative Anstieg in den Varianzen aufgrund der imputierten Werte (Indikator r) ist bei allen Portalen fur alle ermittelten Pfadkoeffizienten gering. Die relative Effizienz der Imputation (Indikator RE) ist fur $m = 5$ im Vergleich zu $m \rightarrow \infty$ bei allen Portalen groer als 0,97, betragt also mehr als 97% (zur Bedeutung der Kennzahlen siehe Anhang A).

Tabelle C.25: t-Statistik zu den gepoolten Pfadkoeffizienten (\bar{Q}) für P1

Pfad	\bar{Q}	\bar{U}	\hat{B}	\hat{T}	t	[ν]	p	r	$\hat{\lambda}$	RE
INFQ → RISK	0,1128	0,0199	0,0035	0,0241	0,7271	130	>0,05	0,2122	0,0137	0,9973
INFQ → VERT	0,4509	0,0106	0,0016	0,0125	4,0301	164	<0,01	0,1846	0,011	0,9978
INFQ → WISS	0,369	0,0104	0,0023	0,0132	3,2077	89	<0,01	0,2684	0,0194	0,9961
PREQ → RISK	0,1819	0,0187	0,0012	0,0202	1,2813	807	>0,05	0,0757	0,0024	0,9995
PREQ → VERT	0,1113	0,0087	0,0018	0,0109	1,0676	104	>0,05	0,2438	0,0168	0,9966
PREQ → WISS	0,2805	0,0083	0,0006	0,009	2,9652	732	<0,01	0,0798	0,0026	0,9995
RISK → LOYA	0,0527	0,0046	0,0022	0,0072	0,6208	31	>0,05	0,5573	0,048	0,9905
TECQ → RISK	-0,1189	0,0168	0,0034	0,0208	-0,8234	106	>0,05	0,2406	0,0165	0,9967
TECQ → VERT	0,1649	0,0115	0,0004	0,012	1,5071	2981	>0,05	0,038	0,0007	0,9999
TECQ → WISS	0,1066	0,01	0,0026	0,0131	0,9317	70	>0,05	0,3133	0,024	0,9952
VERT → LOYA	0,4333	0,0051	0,001	0,0064	5,4282	103	<0,01	0,2447	0,0169	0,9966
VERT → RISK	0,1586	0,0156	0,0113	0,0291	0,9294	18	>0,05	0,8707	0,0715	0,9859
WISS → LOYA	0,4001	0,0041	0,0002	0,0044	6,026	1075	<0,01	0,0649	0,0018	0,9996

Tabelle C.26: t-Statistik zu den gepoolten Pfadkoeffizienten (\bar{Q}) fur P2

Pfad	\bar{Q}	\bar{U}	\bar{B}	\bar{T}	t	[ν]	p	r	$\hat{\lambda}$	RE
INFQ → RISK	0,1606	0,0195	0,0056	0,0262	0,9921	60	>0,05	0,3481	0,0276	0,9945
INFQ → VERT	0,242	0,0089	0,0014	0,0106	2,3549	159	<0,01	0,188	0,0113	0,9977
INFQ → WISS	0,4218	0,011	0,0003	0,0113	3,9616	3915	<0,01	0,033	0,0005	0,9999
PREQ → RISK	-0,1898	0,0184	0,0119	0,0327	-1,0495	20	>0,05	0,7794	0,0655	0,9871
PREQ → VERT	0,2618	0,0097	0,0033	0,0136	2,2409	47	<0,05	0,4067	0,0336	0,9933
PREQ → WISS	0,1588	0,0127	0,0009	0,0139	1,3491	607	>0,05	0,0883	0,0031	0,9994
RISK → LOYA	0,0523	0,0033	0,0002	0,0035	0,8817	831	>0,05	0,0746	0,0023	0,9995
TECQ → RISK	0,0477	0,0175	0,0058	0,0245	0,3047	49	>0,05	0,3987	0,0328	0,9935
TECQ → VERT	0,2822	0,0064	0,0033	0,0104	2,7739	27	<0,01	0,6255	0,0538	0,9894
TECQ → WISS	0,1241	0,0084	0,0031	0,0121	1,1287	42	>0,05	0,4413	0,0371	0,9926
VERT → LOYA	0,3735	0,0065	0,0047	0,0122	3,3857	18	<0,01	0,8599	0,0708	0,986
VERT → RISK	0,1247	0,0218	0,0114	0,0355	0,6616	26	>0,05	0,6281	0,054	0,9893
WISS → LOYA	0,4462	0,0058	0,0025	0,0088	4,7621	34	<0,01	0,515	0,0441	0,9913

Tabelle C.27: t-Statistik zu den gepoolten Pfadkoeffizienten (\bar{Q}) für P3

Pfad	\bar{Q}	\bar{U}	\hat{B}	\hat{T}	t	[ν]	p	r	$\hat{\lambda}$	RE
INFQ → RISK	0,0589	0,0311	0,0174	0,052	0,2584	24	>0,05	0,6711	0,0575	0,9886
INFQ → VERT	0,263	0,0145	0,0005	0,0151	2,1383	2379	<0,05	0,0427	0,0008	0,9998
INFQ → WISS	0,4409	0,0104	0,0011	0,0118	4,0665	310	<0,01	0,1281	0,006	0,9988
PREQ → RISK	-0,1527	0,1286	0,2463	0,4242	-0,2345	8	>0,05	2,298	0,116	0,9773
PREQ → VERT	0,1643	0,02	0,0087	0,0304	0,9423	34	>0,05	0,522	0,0448	0,9911
PREQ → WISS	0,2061	0,0214	0,0027	0,0246	1,3129	238	>0,05	0,1486	0,0077	0,9985
RISK → LOYA	-0,0179	0,0055	0,0022	0,0081	-0,1988	37	>0,05	0,4835	0,0412	0,9918
TECQ → RISK	0,1372	0,0873	0,0536	0,1516	0,3524	22	>0,05	0,737	0,0625	0,9877
TECQ → VERT	0,3239	0,0164	0,0091	0,0274	1,9581	25	>0,05	0,6661	0,0571	0,9887
TECQ → WISS	0,181	0,0208	0,0015	0,0227	1,2021	606	>0,05	0,0884	0,0031	0,9994
VERT → LOYA	0,4852	0,0063	0,0004	0,0067	5,914	806	<0,01	0,0758	0,0024	0,9995
VERT → RISK	0,0609	0,0284	0,0115	0,0422	0,2964	37	>0,05	0,4849	0,0413	0,9918
WISS → LOYA	0,4041	0,0069	0,0003	0,0073	4,7414	1803	<0,01	0,0494	0,0011	0,9998

Tabelle C.28: t-Statistik zu den gepoolten Pfadkoeffizienten (\bar{Q}) fur P4

Pfad	\bar{Q}	\bar{U}	\bar{B}	\bar{T}	t	[ν]	p	r	$\hat{\lambda}$	RE
INFQ → RISK	-0,123	0,0146	0,0032	0,0184	-0,9061	93	>0,05	0,2616	0,0187	0,9963
INFQ → VERT	0,357	0,0083	0,0017	0,0104	3,5031	102	<0,01	0,2456	0,017	0,9966
INFQ → WISS	0,5295	0,0091	0,0007	0,0099	5,3099	540	<0,01	0,0941	0,0035	0,9993
PREQ → RISK	0,0135	0,014	0,0033	0,018	0,1011	84	>0,05	0,2787	0,0204	0,9959
PREQ → VERT	0,2843	0,0081	0,0017	0,0102	2,8132	95	<0,01	0,2571	0,0182	0,9964
PREQ → WISS	0,1765	0,0116	0,0006	0,0124	1,5863	1009	>0,05	0,0672	0,0019	0,9996
RISK → LOYA	0,0148	0,005	0,0035	0,0092	0,1548	19	>0,05	0,8459	0,0699	0,9862
TECQ → RISK	0,2023	0,0121	0,0037	0,0165	1,5736	55	>0,05	0,3684	0,0297	0,9941
TECQ → VERT	0,1813	0,0069	0,0006	0,0077	2,066	409	<0,05	0,1097	0,0046	0,9991
TECQ → WISS	0,0334	0,0071	0,0007	0,008	0,3749	356	>0,05	0,1184	0,0053	0,9989
VERT → LOYA	0,4038	0,0076	0,0033	0,0116	3,75	33	<0,01	0,526	0,0451	0,9911
VERT → RISK	0,4208	0,0127	0,0028	0,016	3,3265	92	<0,01	0,2624	0,0187	0,9963
WISS → LOYA	0,3681	0,0056	0,0014	0,0073	4,3018	72	<0,01	0,3066	0,0233	0,9954

C.5 Bestimmtheitsmaß R^2

Die gepoolten R^2 -Werte (\bar{Q}) und die zugehörigen Standardabweichungen ($\sqrt{\hat{B}}$) sind in den Tabellen C.29, C.30, C.31 und C.32 aufgeführt. Die bei allen Portalen hohen R^2 -Werte der Konstrukte WISS (Wissen), VERT (Vertrauen) und LOYA (Loyalität) zeigen, dass ein beträchtlicher Teil ihrer Varianz durch die anderen Konstrukte des Pfadmodells erklärt wird. Ausnahme ist RISK (Risiko), was darauf hindeutet, dass neben VERT noch andere hier nicht berücksichtigte Faktoren, wie z. B. die persönliche Einstellung, das empfundene Risiko beeinflussen.

Tabelle C.29: Gepoolte R^2 -Werte (\bar{Q}) für P1

Konstrukt	\bar{Q}	$\sqrt{\hat{B}}$
LOYA	0,5871	0,0243
RISK	0,1093	0,0526
VERT	0,4483	0,0184
WISS	0,4732	0,0209

Tabelle C.30: Gepoolte R^2 -Werte (\bar{Q}) für P2

Konstrukt	\bar{Q}	$\sqrt{\hat{B}}$
LOYA	0,5586	0,0262
RISK	0,0491	0,0219
VERT	0,4858	0,0476
WISS	0,4078	0,0188

Tabelle C.31: Gepoolte R^2 -Werte (\bar{Q}) für P3

Konstrukt	\bar{Q}	$\sqrt{\hat{B}}$
LOYA	0,6329	0,0219
RISK	0,0832	0,0519
VERT	0,4634	0,0095
WISS	0,5641	0,0141

Tabelle C.32: Gepoolte R^2 -Werte (\bar{Q}) für P4

Konstrukt	\bar{Q}	$\sqrt{\hat{B}}$
LOYA	0,4941	0,0196
RISK	0,2478	0,0363
VERT	0,5197	0,02
WISS	0,4654	0,0366

C.6 Cronbachs α

Die gepoolten α -Werte (\bar{Q}) und die zugehörigen Standardabweichungen ($\sqrt{\hat{B}}$) sind in den Tabellen C.33, C.34, C.35 und C.36 aufgeführt. Die α -Werte sind bei allen Portalen recht hoch ($>0,6$) und deuten auf eine gute Reliabilität der Konstrukte hin. Ausnahme bei P2–P4 ist VERT (Vertrauen) mit Werten von knapp unter der für Untersuchungen mit explorativem Charakter noch akzeptablen Untergrenze von 0,6. Für eine bessere Reliabilität wäre eine Reformulierung der Fragen VERT1 und VERT2 bei zukünftigen Untersuchungen zu erwägen.

Tabelle C.33: Gepoolte α -Werte (\bar{Q}) für P1

Konstrukt	\bar{Q}	\sqrt{B}
INFQ	0,8473	0,0047
LOYA	0,9193	0,0172
PREQ	0,8313	0,0017
RISK	0,7314	0,0865
TECQ	0,7644	0,0137
VERT	0,6241	0,0472
WISS	0,8529	0,0105

Tabelle C.34: Gepoolte α -Werte (\bar{Q}) für P2

Konstrukt	\bar{Q}	\sqrt{B}
INFQ	0,8292	0,0044
LOYA	0,7841	0,0044
PREQ	0,8668	0,0035
RISK	0,6597	0,0896
TECQ	0,7812	0,0036
VERT	0,5683	0,0372
WISS	0,7949	0,0066

Tabelle C.35: Gepoolte α -Werte (\bar{Q}) für P3

Konstrukt	\bar{Q}	\sqrt{B}
INFQ	0,8566	0,008
LOYA	0,8464	0,0108
PREQ	0,8751	0,0039
RISK	0,6071	0,166
TECQ	0,8292	0,0082
VERT	0,5679	0,0838
WISS	0,6796	0,0146

Tabelle C.36: Gepoolte α -Werte (\bar{Q}) für P4

Konstrukt	\bar{Q}	\sqrt{B}
INFQ	0,8327	0,0041
LOYA	0,9006	0,0063
PREQ	0,8267	0,0015
RISK	0,6793	0,1199
TECQ	0,7151	0,0072
VERT	0,5556	0,0628
WISS	0,8411	0,0094