

Abbildungsverzeichnis.....	XV
Tabellenverzeichnis.....	XIX
Symbolverzeichnis.....	XXI
Abkürzungsverzeichnis.....	XXIII
<b>1 Einführung.....</b>	<b>1</b>
1.1 Motivation .....	1
1.2 Gang der Arbeit .....	4
<b>2 Marketingforschung in internetbasierten Märkten .....</b>	<b>7</b>
2.1 Grundlagen elektronischer Märkte .....	7
2.1.1 Definition und Eigenschaften elektronischer Märkte .....	7
2.1.2 Architekturmodell elektronischer Marktsysteme .....	14
2.2 Architektur internetbasierter Marktsysteme .....	23
2.2.1 Kommunikationsdienste des Internet.....	23
2.2.2 Architektur internetbasierter Marktapplikationen.....	27
2.3 Implikationen internetbasierter Marktsysteme für die Marketingforschung .....	30
2.3.1 Methodologische Grundlagen der Marketingforschung .....	30
2.3.2 Instrumente der Datenerhebung .....	34
2.3.2.1 Erfassung des erhebungstechnischen Instrumentariums .....	34
2.3.2.2 Instrumente zur indirekten Beobachtung des Konsumentenverhaltens .....	39
2.3.2.3 Rechtliche und ethische Rahmenbedingungen.....	53
2.3.3 Instrumente der Datenanalyse .....	55
2.3.3.1 Ableitung eines generischen Architekturkonzepts .....	55
2.3.3.2 Deckungsgrad verfügbarer Datenanalysesysteme .....	58
<b>3 Data Mining und Knowledge Discovery in Databases als innovative Datenanalysekonzepte .....</b>	<b>61</b>
3.1 Begriffsdefinition und Abgrenzung.....	61
3.1.1 Positionsbestimmung Data Mining und Knowledge Discovery in Databases .....	61
3.1.2 Definition des Data Mining.....	63
3.1.3 Definition des Knowledge Discovery in Databases (KDD).....	65
3.1.4 Abgrenzung der Konzepte.....	67
3.2 Prozeßorientierte Darstellung des KDD-Konzepts.....	70
3.2.1 Der KDD-Prozeß.....	70
3.2.2 Selektion und Extraktion.....	73
3.2.3 Vorbereitung und Transformation.....	74
3.2.3.1 Datenbereinigung .....	74
3.2.3.2 Datencodierung .....	77
3.2.3.3 Transformation .....	79

3.2.4 Mustererkennung .....	79
3.2.4.1 Konfiguration des Algorithmus.....	79
3.2.4.2 Anwendung des konfigurierten Algorithmus.....	81
3.2.5 Evaluation .....	87
3.2.6 Präsentation.....	91
3.2.7 Interpretation der Ergebnisse .....	93
3.3 Methoden des Data Mining .....	95
3.3.1 Systematisierung der Data Mining-Methoden .....	95
3.3.2 Darstellung ausgewählter Data Mining-Methoden .....	103
3.3.2.1 Assoziationsanalyse .....	103
3.3.2.2 Sequenzanalyse .....	107
3.3.2.3 Entscheidungsbauminduktion .....	109
3.3.2.4 Clusteranalyse .....	114
3.3.2.5 Künstliche Neuronale Netze.....	121
3.3.2.6 On-Line Analytical Processing (OLAP).....	125
3.4 Fazit.....	129
<b>4 Web Log Mining als Datenanalysekonzept für internetbasierte Märkte.....</b>	<b>131</b>
4.1 Abgrenzung und Begriffsdefinition.....	131
4.2 Prozeßorientierte Darstellung des WLM-Konzepts .....	133
4.2.1 Datengenerierung .....	133
4.2.2 Selektion und Extraktion.....	134
4.2.3 Vorbereitung und Transformation.....	138
4.2.3.1 Bereinigung .....	138
4.2.3.2 Transaktionsableitung .....	138
4.2.3.3 Datencodierung .....	143
4.2.3.4 Transformation.....	145
4.2.4 Mustererkennung .....	146
4.2.4.1 Vorbemerkung.....	146
4.2.4.2 Assoziationsanalyse .....	146
4.2.4.3 Sequenzanalyse .....	149
4.2.4.4 Clusteranalyse .....	150
4.2.4.5 Entscheidungsbauminduktion .....	154
4.2.5 Evaluation .....	156
4.2.6 Präsentation.....	159
4.2.7 Interpretation.....	163
4.2.8 Fazit.....	163
4.3 Systemanalyse.....	164
4.3.1 Übersicht .....	164
4.3.2 Webminer.....	165
4.3.3 WebLogMiner .....	167
4.3.4 SpeedTracer .....	170

4.3.5 Web Utilization Miner .....	173
4.3.6 Ergebnisse der Systemanalyse.....	175
<b>5 Entwicklung des WLM-Systems AutoGen.....</b>	<b>177</b>
5.1 Softwaretechnische Grundlagen .....	177
5.2 Anforderungsanalyse .....	179
5.3 Entwurf.....	182
5.3.1 Systemarchitektur.....	182
5.3.2 Systemkomponenten .....	184
5.3.2.1 Managementkomponenten .....	184
5.3.2.1.1 Konfigurationsdatenbank.....	184
5.3.2.1.2 Management-Frontend.....	186
5.3.2.2 Komponenten des Web Log Warehouse .....	186
5.3.2.2.1 Monitor und Extraktor .....	186
5.3.2.2.2 Integrator.....	188
5.3.2.2.3 Web Log Warehouse-Datenbank .....	190
5.3.2.3 Web Log Mining-Komponenten .....	192
5.3.2.3.1 Web Log Mining-Datenbank .....	192
5.3.2.3.2 Web Log Mining-Server .....	196
5.3.2.3.3 Web Log Mining-Frontend .....	199
5.4 Implementierung.....	201
5.4.1 Basistechnologien .....	201
5.4.1.1 Programmiersprachen und Entwicklungsumgebungen .....	201
5.4.1.2 Kommunikationsprotokolle.....	202
5.4.1.3 Datenbanktechnologie .....	204
5.4.2 Mustererkennung .....	205
5.4.3 Komponentenverteilung.....	207
5.4.4 Systemablauf.....	209
<b>6 Evaluation des AutoGen-Systems .....</b>	<b>217</b>
6.1 Datengrundlage.....	217
6.2 Parameterbelegung .....	218
6.3 Systemanwendung .....	222
6.3.1 Hypothesengenerierung.....	222
6.3.2 Hypothesenprüfung .....	232
6.3.3 Laufzeitverhalten.....	236
6.4 Zusammenfassende Beurteilung des AutoGen-Systems .....	237
<b>7 Zusammenfassung und Ausblick .....</b>	<b>241</b>
<b>Literaturverzeichnis.....</b>	<b>247</b>

## Abbildungsverzeichnis

Abb. 1:	Struktur der Arbeit im Überblick .....	5
Abb. 2:	Phasen der Markttransaktion .....	7
Abb. 3:	Typen elektronischer Märkte .....	12
Abb. 4:	Kommunikationsmodell nach SHANNON und WEAVER .....	14
Abb. 5:	OSI-Referenzmodell .....	16
Abb. 6:	Klassifikation von Marktdiensten .....	18
Abb. 7:	Client/Server-Modell als Strukturierungsrahmen für Marktapplikationen .....	21
Abb. 8:	Architekturmodell elektronischer Marktsysteme .....	22
Abb. 9:	Client/Server-Modell des WWW .....	25
Abb. 10:	Zusammenfassende Darstellung der TCP/IP-Protokollfamilie .....	26
Abb. 11:	Dokumentenzentrierte Architektur internetbasierter Marktapplikationen .....	28
Abb. 12:	Java-basierte Architektur internetbasierter Marktapplikationen .....	29
Abb. 13:	Zusammenhang zwischen Marketingtheorie, Marketingentscheidung, Marketingforschung und Marktverhalten .....	31
Abb. 14:	Darstellung des Induktionsprinzips .....	33
Abb. 15:	Darstellung des Induktionsproblems .....	33
Abb. 16:	Beobachtungsformen in internetbasierten Märkten .....	37
Abb. 17:	Konzept der serverbasierten Protokollaufzeichnung .....	40
Abb. 18:	Auszug eines typischen Zugriffsprotokolls im CLF-Format .....	41
Abb. 19:	Attribute der Protokolldateien .....	41
Abb. 20:	Stern-Schema der Protokolldaten .....	46
Abb. 21:	Einsatz eines Proxy-Server zum Zwischenspeichern von Ressourcen .....	49
Abb. 22:	Generisches Architekturkonzept eines Datenanalysesystems zur Marketingforschung .....	56
Abb. 23:	Abgrenzung von Data Mining und KDD .....	70
Abb. 24:	Grafische Darstellung des KDD-Prozesses .....	72
Abb. 25:	Darstellung des Bottom Up- und Top Down-Verfahrens .....	82
Abb. 26:	Darstellung eines Suchgrafen .....	83
Abb. 27:	Darstellung der Suchgrafen für die erschöpfende und heuristische Suche .....	85
Abb. 28:	Aufgaben der Musterevaluation .....	91
Abb. 29:	Medientypen .....	91
Abb. 30:	Klassifikationsverfahren des Data Mining .....	98
Abb. 31:	Segmentierungsverfahren des Data Mining .....	100
Abb. 32:	Darstellung der Sequenzanalyse .....	108
Abb. 33:	Darstellung eines binären Entscheidungsbaums .....	110
Abb. 34:	Überblick ausgewählter Cluster-Verfahren .....	118
Abb. 35:	Dendrogramm zur Darstellung der Clusterbildung .....	119
Abb. 36:	Grundmodell eines künstlichen Neurons .....	122
Abb. 37:	Aufbau eines Künstlichen Neuronalen Netzes .....	123
Abb. 38:	Dimensionshierarchie mit multiplen Konsolidierungspfaden .....	127
Abb. 39:	Grafische Darstellung der Dicing-Technik .....	128
Abb. 40:	Grafische Darstellung der Slicing-Technik .....	128
Abb. 41:	Taxonomie des Web Mining .....	131
Abb. 42:	Web Log Mining-Prozeß .....	133
Abb. 43:	Verweisintegration komplementärer Informationsangebote .....	148

Abb. 44:	Navigationspfad einer Transaktion.....	149
Abb. 45:	Erfassung der Grund- und Interaktionsdaten als Segmentierungskriterien .....	152
Abb. 46:	Segmentierungsbeispiel auf der Basis verhaltensorientierter und technografischer Kriterien .....	153
Abb. 47:	Kriterien zur Klassifikation von Transaktionen .....	155
Abb. 48:	Klassifikation von Transaktionen nach dem Kriterium der Verweildauer .....	156
Abb. 49:	Objektähnlichkeitssuche zur Musterevaluation.....	157
Abb. 50:	Grafische Präsentationsformen für Assoziationsregeln.....	159
Abb. 51:	Grafische Präsentation von Pfadmustern.....	160
Abb. 52:	Grafische Präsentation von Segmenten mit dem Verfahren der parallelen Koordinaten .....	161
Abb. 53:	Grafische Präsentation eines Entscheidungsbaums durch Anwendung des hyperbolischen Geometriemodells .....	162
Abb. 54:	Prozeßmodell des Webminer-Systems .....	165
Abb. 55:	Prozeßmodell des Systems WebLogMiner.....	167
Abb. 56:	Beispiel einer Anwendung der Dicing-Technik (WebLogMiner).....	169
Abb. 57:	Prozeßmodell des SpeedTracer-Systems.....	170
Abb. 58:	Gruppenbericht von SpeedTracer .....	171
Abb. 59:	Pfadbericht von SpeedTracer .....	172
Abb. 60:	Benutzerbericht von SpeedTracer .....	172
Abb. 61:	Prozeßmodell des Systems Web Utilization Miner .....	173
Abb. 62:	MINT-Abfrage (Web Utilization Miner) .....	174
Abb. 63:	Visualisierung der entdeckten Pfadmuster (Web Utilization Miner) .....	175
Abb. 64:	Anwendungsfalldiagramm des Systems AutoGen .....	179
Abb. 65:	Komponentenarchitektur des Systems AutoGen.....	183
Abb. 66:	ER-Modell der Konfigurationsdatenbank .....	185
Abb. 67:	Funktionen des Management-Frontend .....	186
Abb. 68:	Prozeßmodell des Monitors.....	187
Abb. 69:	Prozeßmodell des Extraktors.....	188
Abb. 70:	Prozeßmodell des Integrators .....	189
Abb. 71:	ER-Modell der WLW-Datenbank .....	191
Abb. 72:	ER-Modell der Hypothesenbank .....	193
Abb. 73:	ER-Modell der Musterbank.....	195
Abb. 74:	Prozeßmodell des WLM-Server .....	198
Abb. 75:	Funktionen des WLM-Frontend .....	199
Abb. 76:	Visualisierung der zeitlichen Entwicklung einer Hypothese (Hypothesentrajektorie).....	200
Abb. 77:	RMI-Schichtenmodell .....	203
Abb. 78:	Plan Designer als interaktives Entwurfswerkzeug .....	206
Abb. 79:	Komponentenverteilungsmodell von AutoGen.....	208
Abb. 80:	Management-Frontend von AutoGen.....	210
Abb. 81:	E-Mail-Bericht.....	211
Abb. 82:	Hauptmenü des WLM-Frontend.....	211
Abb. 83:	Präsentation der Musterbank .....	212
Abb. 84:	Verwaltung der Hypothesenklassen .....	213
Abb. 85:	Hypothesentransfer .....	213
Abb. 86:	Präsentation der Hypothesenbank .....	214

Abb. 87:	AutoGen Hypothesenbericht .....	215
Abb. 88:	Veranstaltungsinformationssystem NightZap .....	217
Abb. 89:	Häufigkeitsverteilung der Referenzdauer .....	219
Abb. 90:	Regelanzahl für unterschiedliche Constraint-Belegungen .....	221
Abb. 91:	Exploration der Musterbank nach dem ersten Analysezyklus .....	223
Abb. 92:	Anzahl der abgeleiteten Transaktionen und der entdeckten Regeln .....	224
Abb. 93:	Browserbasierte Anzeige der Rubrik <i>Top Ten</i> .....	226
Abb. 94:	Hierarchische Verbundbeziehung .....	226
Abb. 95:	Zeitliche Entwicklung der hierarchischen Verbundbeziehung .....	227
Abb. 96:	Transfer der Suchverhaltenshypothese in die Hypothesenbank .....	228
Abb. 97:	Grafische Darstellung der informatorischen Verbundbeziehungen .....	229
Abb. 98:	Zeitliche Entwicklung der Konfidenz der informatorischen Verbundbeziehungen (Regeln 2–5) .....	230
Abb. 99:	Zeitliche Entwicklung des Support der informatorischen Verbundbeziehungen (Regeln 2–5) .....	231
Abb. 100:	Transfer einer informatorischen Verbundbeziehung in die Hypothesenbank .....	232
Abb. 101:	Trajektorieausschnitt der Suchverhaltenshypothese .....	233
Abb. 102:	Zeitliche Entwicklung der Suchverhaltenshypothese .....	234
Abb. 103:	Zeitliche Entwicklung der Konfidenz der Informationsverhaltenshypothesen ...	234
Abb. 104:	Zeitliche Entwicklung des Support der Informationsverhaltenshypothesen .....	235
Abb. 105:	Analysedauer in Abhängigkeit von der Anzahl der Elementaroperationen .....	237
Abb. 106:	Strukturmodell einer adaptiven Marktapplikation .....	245

## Tabellenverzeichnis

Tab. 1:	Organisationstypen von Märkten.....	8
Tab. 2:	Anwendungsorientierte Dienste und Protokolle des Internet.....	24
Tab. 3:	Datenklassifikation des Database Marketing-Konzepts.....	45
Tab. 4:	Systematisierung der Attribute der Protokollkomponente anhand des Paradigmas des Informationsverhaltens.....	47
Tab. 5:	Semiotische Begriffsabgrenzung.....	68
Tab. 6:	Phasen des KDD-Prozesses.....	71
Tab. 7:	Kombinationsmöglichkeiten beim Vergleich binärer Attribute.....	116
Tab. 8:	Ausgewählte Ähnlichkeitsmaße für binäre Attribute.....	117
Tab. 9:	Diskrepanzen zwischen transaktionsorientierter und entscheidungsorientierter Datenverarbeitung.....	126
Tab. 10:	Abruf grafischer Elemente.....	135
Tab. 11:	Fehlerhafte HTTP-Anforderung.....	135
Tab. 12:	Rückübertragung von Daten an den WWW-Server.....	136
Tab. 13:	Aufspaltung zusammengesetzter Attribute in Einzelattribute.....	137
Tab. 14:	Exemplarische Protokolldaten.....	140
Tab. 15:	Transaktionsableitung auf der Basis der Browserdaten und Zeit ( $\Delta t^{\text{krit}} = 30$ Min.).....	141
Tab. 16:	Pfadanalytische Transaktionsableitung.....	142
Tab. 17:	Verdichtung von Transaktionen.....	143
Tab. 18:	Codierung der HTTP-Anforderung.....	144
Tab. 19:	Ermittlung der Referenzdauer.....	144
Tab. 20:	Übersicht der Web Log Mining-Systeme.....	164
Tab. 21:	Dimensionen und Hierarchien des Systems WebLogMiner.....	168
Tab. 22:	Ergebnisse der Systemanalyse.....	175
Tab. 23:	Relationen der WLM-Datenbank.....	196
Tab. 24:	Methodenbank des Data Mining-Systems K.wiz.....	206
Tab. 25:	Rubriken von NightZap.....	218
Tab. 26:	Werte der exponentiellen Verteilungsfunktion.....	220
Tab. 27:	Verteilung der Regelgröße für unterschiedliche Constraint-Belegungen.....	222
Tab. 28:	Übersicht der entdeckten Zusammenhänge nach Häufigkeit.....	225
Tab. 29:	Ergebnisse der Hypothesenprüfung.....	233