

Arbeitsbericht Nr. 17/2003

Hrsg.: Matthias Schumann

Christian Kaspar / Svenja Hagenhoff

**Individualität und Produktindividualisierung -
Kundenprofile für die Personalisierung von digitalen Produkten**

Inhaltsverzeichnis

Abbildungsverzeichnis	IV
Abkürzungsverzeichnis	V
1 Einleitung	1
2 Die ökonomische Rolle der menschlichen Individualität.....	1
2.1 Auslöser von Individualisierungsprozessen	2
2.1.1 Gesellschaftliche Liberalisierung	3
2.1.2 Entscheidungskomplexität und beschränkte menschliche Rationalität	4
2.1.3 Digitale Informations- und Kommunikationstechnik.....	6
2.2 Individuell orientierte betriebliche Handlungsmöglichkeiten.....	7
2.2.1 Wettbewerbsstrategien und Individualität.....	7
2.2.2 Wissensintensivierung und Individualität.....	9
2.2.3 Geschäftsmodelle und Individualität.....	10
2.3 Zwischenfazit: Individualität und Individualisierung.....	11
3 Individualität und individuelle Identifikation	12
3.1 Registrierung und Präferenzprofile.....	13
3.1.1 Erhebung von Kundenprofilen durch Befragung	13
3.1.2 Identifikation des Kunden über registrierte Profile.....	15
3.1.3 Bewertung von freiwilligen Benutzerprofilen zur individuellen Kundenidentifikation	16
3.2 Web Usage Mining und Nutzungsprofile	17
3.2.1 Data Mining und Knowledge Discovery in Databases.....	18
3.2.2 Datenvorbereitung und -transformation von Webserver Protokolldateien	20
3.2.3 Mustererkennung in Webserver Protokolldateien.....	22
3.2.3.1 Assoziationsanalyse	22
3.2.3.2 Clusteranalyse	23
3.2.4 Bewertung des Web Usage Mining für die individuelle Nutzeridentifikation	23
3.3 Rechtliche Grundlagen der Nutzung personenbezogener Daten	24
4 Produktindividualisierung	25

4.1 Individuelle Empfehlungssysteme	27
4.1.1 Individualisierung durch explizite Nutzerangaben	28
4.1.2 Individualisierung durch Collaborative Filtering	30
4.1.3 Individualisierung durch beobachtungsbasiertes Filtern	33
4.1.3.1 Profilaggregation im Rahmen eines beobachtungsbasierten Empfehlungssystems	34
4.1.3.2 Ermittlung einer individuellen Empfehlung im Rahmen eines beobachtungsbasierten Empfehlungssystems	36
4.1.4 Bewertung der dargestellten Verfahren	37
4.2 Praktische Anwendungen von Empfehlungssystemen im Rahmen der Produktindividualisierung	39
4.2.1 Praxisbeispiele für Individualisierungsmöglichkeiten im Internet	39
4.2.2 Bewertung aktueller Individualisierungsformen im Internet	41
5 Zusammenfassung	43
Literaturverzeichnis	44

Abbildungsverzeichnis

Abbildung 1: Rahmenbedingungen der Individualisierung	2
Abbildung 2: Freiheit und Individualität	4
Abbildung 3: Individuelle Strategieorientierungen	7
Abbildung 4: Synthese - Rahmen- und Gestaltungsbedingungen der Individualisierung	11
Abbildung 5: Qualität von Benutzerprofilen	12
Abbildung 6: Profilregistrierung bei Amazon	14
Abbildung 7: P3P Kommunikationsarchitektur	16
Abbildung 8: Struktur eines Server Logfiles	18
Abbildung 9: Anwendungsbereiche des Data Mining im Internet	19
Abbildung 10: KDD	20
Abbildung 11: Datenbereinigung im Zugriffsprotokoll	21
Abbildung 12: Binärcodierung zur Ähnlichkeitsbestimmung	23
Abbildung 13: Individualisierungs-Regel-System	27
Abbildung 14: Verfahren zur Produktindividualisierung	27
Abbildung 15: Ähnlichkeitsbestimmung durch eigenschaftsbasiertes Filtern	29
Abbildung 16: Datenmatrix der Ratingpräferenzen beim Collaborative Filtering	30
Abbildung 17: Content-boosted Collaborative Filtering	32
Abbildung 18: Konzept eines beobachtungsbasierten Empfehlungssystems.....	34
Abbildung 19: Profilaggregation durch Transaktionsclustering	35
Abbildung 20: Profilaggregation durch Pageviewclustering	36
Abbildung 21: Bewertung verschiedener Personalisierungsmethoden.....	38
Abbildung 22: Praxisbeispiel 1 - Mein Yahoo!.....	39
Abbildung 23: Praxisbeispiel 2 - Nike-ID	40
Abbildung 24: Individualisierungsformen und Anwendungsbeispiele	41
Abbildung 25: Individualisierungsmöglichkeiten für Medienprodukte	43

Abkürzungsverzeichnis

BDSG	Bundesdatenschutzgesetz
CF	Collaborative Filtering
HTTP	Hypertext Transfer Protocol
IP	Internet Protocol
IuK-Technik	Informations- und Kommunikationstechnik
KDD	Knowledge Discovery in Databases
OPS	Open Profiling Standard
P3P	Platform for Privacy Preferences Protocol
RFC	Request for Comment
TDDSG	Teledienste Datenschutzgesetz
TDG	Gesetz über die Nutzung von Telediensten
URL	Uniform Resource Locator
W3C	World Wide Web Consortium

1 Einleitung

Die Frage der optimalen Sortenvarietät bildet von je her ein zentrales Problem der strategischen Planung. Im Zuge der Verbreitung des Internets hat die Varietätsplanung jedoch eine neue Qualität erfahren. Zum einen repräsentiert das Internet einen zusätzlichen Distributionskanal für Informationsgüter aller Art mit der Möglichkeit eines interaktiven Dialogs zwischen Kunden und Produkthanbietern. Zum anderen ermöglicht die Verbreitung digitaler Informations- und Kommunikationstechnik in Verbindung mit dem Internet zunehmend die nahtlose und effiziente Integration von Interaktions- und Feedbackdaten aus diesem Dialog in betriebliche Leistungsprozesse. Die Folge ist eine potenziell höhere Transparenz bzgl. des Verhaltens und der Bedürfnisse eines Kunden bei gleichzeitiger verbesserter Effizienz der Integration kundenbezogener Daten.

Aus diesen Potenzialen der Internettechnologien leiten sich eine Vielzahl von theoretischen und jüngst auch ersten praktischen Konzepten ab, mit dem Ziel, diese Potenziale im Sinne einer kundenindividuellen Varietät auszuschöpfen. Im Folgenden soll der derzeitige Stand der betriebswirtschaftlichen Forschung in Bezug auf die Generierung und Integration von Kundenprofilen für eine Individualisierung der Leistungsbeziehung im Internet vorgestellt werden. Dabei werden im Rahmen einer Einführung zunächst thematische Grundlagen in der Literatur in Bezug auf die Bedeutung der Individualisierung abgegrenzt. Daran anschließend erfolgt ein Überblick über verschiedene Verfahren zur Ableitung von individuellen Kundenprofilen mit einem kurzen Exkurs über die aktuelle Rechtslage bzgl. des Datenschutzes. Im vierten Kapitel werden Instrumente zur Produktindividualisierung vorgestellt, offene Fragen in der Literatur in diesem Zusammenhang identifiziert und Anregungen für eine weiterführende Forschung geliefert.

2 Die ökonomische Rolle der menschlichen Individualität

Die klassische Mikroökonomie liberalistischer Prägung unterstellt eine eingeschränkt individualistische Gesellschaftskonzeption, in der Konsumenten ihren individuellen Nutzen im Rahmen der Befriedigung ihrer Bedürfnisse durch den Konsum von Gütern maximieren.¹ Das Marktangebot entsteht als gewinnoptimale Mengenanpassung der Hersteller ihres Produktionsvolumens an den gegebenen Marktpreis. Neben den modellimmanenten Vereinfachungen wie vollkommene Preistransparenz und atomistische Angebots- und Nachfragestrukturen wird in diesem Grundmodell insbesondere die Granularität individueller Konsumentenpräferenzen verkürzt dargestellt, indem von einer weitgehenden Homogenität von Güterangebot und -nachfrage ausgegangen wird. Individuelle Präferenzordnungen der Modellwelt unterscheiden lediglich zwischen Güterarten zur Befriedigung desselben Mangels, nicht aber bspw. zwischen verschiedenen Gütern derselben Güterart, jedoch unterschiedlicher Anbieter. Insofern werden aktuelle soziale und ökonomische Entwicklungstendenzen wie das

¹ Schumann/Meyer/Ströbele (1999), S. 9

wachsende Güterüberangebot und die daraus resultierende Notwendigkeit zur Differenzierung des Warenangebots gegenüber Konkurrenten oder bspw. die zunehmende Hedonisierung und das Wohlstandsbedingte Verblenden des traditionellen Gesellschafts- und Rollenbildes in ihren Auswirkungen dieser Entwicklungen auf Angebot und Nachfrage ausgeklammert.

Aktuelle Werbeslogans wie "Ich will so bleiben wie ich bin!" oder "...weil ich es mir wert bin!" deuten auf den ökonomisch erlöswirksam abschöpfbaren Trend einer Marktorientierung in Richtung der individuellen Besonderheiten und Bedürfnisse des einzelnen Menschen hin. Im folgenden Abschnitt soll deshalb der Charakter und die Bedeutung solcher individuellen Spezifika des Menschen – in seinen verschiedenen ökonomischen Rollen, als Mitglied der Gesellschaft, als Mitarbeiter im Rahmen der Leistungsproduktion und als Konsument – untersucht werden. Prozesse der Individualisierung können dabei in zweierlei Hinsicht interpretiert werden: zum einen als exogene Trends aus Gesellschaft, Technologie oder menschlicher biologischer Natur, zum anderen als endogene Variablen der ökonomischen Gestaltung von betrieblichen Prozessen oder Produkten. Im Folgenden werden exogene Trends und endogene Gestaltungsmöglichkeiten kurz getrennt voneinander vorgestellt.

2.1 Auslöser von Individualisierungsprozessen

Wie in Abbildung 1 dargestellt, ist die Bedeutung und die Rolle der Individualität des Menschen im heutigen Gesellschaftsbild durch drei Bestimmgrößen geprägt: die gesellschaftshistorische Liberalisierung traditioneller Gemeinschaftsformen durch formalisierte, leistungsorientierte Regelwerke (①); die zunehmende Bedeutung individueller Lernerfahrung zur Handhabung komplexer Entscheidungsprobleme unter der Bedingung beschränkter menschlicher Rationalität (②); die gleichzeitige Informationsvielfalt und potenzielle Nutzertransparenz durch die Entwicklungen im Rahmen digitaler Informations- und Kommunikationstechnik (③). Diese Bestimmgrößen werden im Folgenden kurz erläutert.

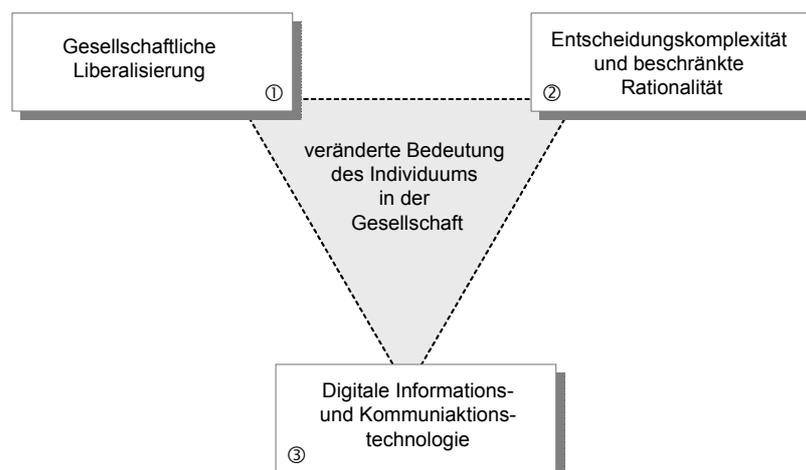


Abbildung 1: Rahmenbedingungen der Individualisierung

2.1.1 Gesellschaftliche Liberalisierung

Erstmalig formulierte Ulrich Beck in den 80er Jahren die These der gesellschaftsweiten Individualisierung vor dem Hintergrund der Prozesse des HerauslöSENS des Individuums aus Schicht- und Klassenmilieus und aus traditionell-industriegesellschaftlichen Formen der Vergesellschaftung und Vergemeinschaftung.² Diese gesellschaftlichen Individualisierungsprozesse können dabei in zweierlei Hinsicht interpretiert werden:³ zum einen auf der Ebene kultureller Codes, als voranschreitende End-traditionalisierung des Gemeinschaftswesen bspw. durch üppigere Wahlmöglichkeiten im Bereich von Massenmedien und Konsum und durch das Schwinden verbindlicher Modelle der Lebensführung; zum anderen auf der sozialstruktureller Ebene, als Gewinn an individueller Freiheit durch die Ausweitung politischer und sozialer Rechte, der Steigerung des materiellen Wohlstands und der Differenzierung und Pluralisierung von Wertorientierungen und Lebensstilen.⁴

Für die Verortung der Rolle des Individuums in der Gesellschaft spielt insbesondere die letztgenannte Forschungstradition eine zentrale Rolle, durch ihre Relativierung zwischen individuellen Lebensläufen und sozialgeschichtlich geprägten Lebensstilen bzw. sozialstrukturellen Milieus. Dieser Denkweise entsprechend kann – wie in Abbildung 2 dargestellt – die heutige Form der begrifflichen Bedeutung und der gesellschaftlichen Rolle der Individualität sozialgeschichtlich als Ergebnis der historischen Herausbildung und Entwicklung von Ordnung- und Sicherheitsmechanismen im Rahmen des gemeinschaftlichen Zusammenlebens hergeleitet werden und ist eng geknüpft an die Entfaltung der individuellen Freiheit im gemeinschaftlichen Gefüge.

² Meier (2003) und Berger (1997), S. 82

³ Vgl. Berger (1997), S. 81

⁴ Vgl. auch Wohlrab-Sahr (1997), S. 24

Das mittelalterliche Ordnungsgefüge im häuslichen wie auch staatlichen Rahmen ist geprägt durch das Herrschaftsprinzip von Huld und Treue, in dem Individualität lediglich als Typologisierung in Form kleinstmöglicher Anteilschaft innerhalb des Gemeinschaftsgefüges auftritt. Eine Thematisierung des Selbst im Sinne einer Seinsdefinition eines Menschen außerhalb solcher typologischer Mitgliedschaften besitzt ihre Wurzeln im neuzeitlichen Staatsverständnis absolutistischer Prägung, durch die Herausbildung von bürgerlichen Freiheiten im Zuge der Aufhebung des Beziehungsgefüges zwischen Herr und Gefolge zu Gunsten nationalstaatlichem Gehorsam. Der Eigennutzen, als wie angesprochen auch heute noch dominierendes Individualprinzip, ist Schöpfung des absolutistischen Rechtsverständnisses, als einerseits verhaltenssteuernde Instanz für das Individuum gegenüber der nationalstaatlichen Leistungskontrolle, andererseits als Schutzbedürfnis der persönlichen Freiheit gegenüber gesellschaftlicher Anarchie durch divergierenden Handlungsmotiven der Mitbürger. Die im 18. Jahrhundert von Adam Smith postulierte von staatlichem Einfluss liberalisierte Ökonomie unterstreicht ökonomisch die Dominanz des individuellen Gewinnstrebens zur Maximierung des Volkseinkommens gegenüber staatlicher Planlenkung, erfordert jedoch gleichzeitig einerseits stabile Rahmenbedingungen für den Leistungsaustausch, andererseits die Sicherung des individuellen Wohlstandes gegenüber negativen Schwankungen im un gelenkten Wirtschaften. Insofern schließen sich Wohlstandssicherung und vollständige individuelle Freiheit (im Sinne von Hayeks als das Fehlen jeglicher handlungsbeschränkender Zwänge) aus.

Abbildung 2: Freiheit und Individualität⁵

Individualität kann also verstanden werden zum einen als Originalität im Sinne des einleitenden Begriffsverständnisses, andererseits aber auch als die Soziabilität dieser Originalität im zwangsläufig verhaltensbeschränkenden Gemeinwesen.⁶ Die derzeit stabilen Verhältnisse durch nationale Wohlfahrt und individueller Rechtssicherheit reduzieren jedoch die Eingriffe gemeinschaftlicher Reglementierungen in die individuellen Freiheiten auf ein im Vergleich zum mittelalterlichen Feudalwesen minimiertes Mindestmaß. Dadurch entsteht die Reproduktion einer individuellen Differenzierung von Konsumansprüchen als eine gesellschaftliche Massenkultur.⁷

2.1.2 Entscheidungskomplexität und beschränkte menschliche Rationalität

Die Schaffung eines stabilen Marktgleichgewichts im Idealbild nach liberalistischem Ökonomieverständnis erfordert eine vollständige Information von Anbietern und Nachfragern über die am Markt gesetzten Preise. Mit Verweisen auf Max Weber beschreibt Kirsch diese implizite Zweckrationalität als das Ergebnis der Prozesse der sozialen Evolution in der abendländischen Gesellschaftsentwicklung.⁸ Vor dem Hintergrund der diversifizierten und differenzierten Angebotsstrukturen eine heute globalisierten Wirtschaft stellt eine solche umfassende Zweckrationalität eine hohe Anforderung an die Denkvermögen im Rahmen der dafür notwendigen Verarbeitung von Informationen an involvierte Akteure. Bereits in den 50er Jahren des letzten Jahrhunderts stellte Simon fest, dass die für eine vollkommene

⁵ Vgl. dazu Sonntag (1999) und Bouillon (1997), S. 38

⁶ Vgl. Sonntag (1999), S. 10

⁷ Vgl. Schweikart (1994), S. 4

⁸ Vgl. Kirsch (1997), S. 619 f.

Rationalität erforderliche Kenntnis und Voraussicht aller möglichen Konsequenzen bei jeder Wahl in der Realität stets lediglich fragmentarisch bleibt.⁹ Die Annahme vollkommener Rationalität stellt deshalb nach Lindblom lediglich ein "synoptisches Ideal"¹⁰ dar, das nicht an bspw. die begrenzten intellektuellen Fähigkeiten des Menschen, sein generell begrenztes Wissen und die Aufwendigkeit vollständiger Situationsanalysen angepasst ist.¹¹ Eine Reihe von Untersuchungen hat sich vor diesem Hintergrund eine präzisierte Analyse menschlicher Entscheidungsprozesse vorgenommen:

Nach Kirsch lässt sich das menschliche Problemlösen im Geiste des radikalen Konstruktivismus¹² als ein Prozess der Informationsverarbeitung beschreiben, in dem der Entscheider durch seine Persönlichkeit in Form seiner im Langzeitgedächtnis gespeicherten Erfahrungen als WahrnehmungsfILTER gegenüber Reizsignalen im kapazitär vergleichsweise beschränkten Kurzzeitgedächtnis fungiert.¹³ Dadurch beschränkt sich das Such- und Auswahlverhalten zwischen Entscheidungsvariablen im Rahmen einer Problemstellung tendenziell in dem Rahmen erfahrungsgeprägter Muster, innerhalb derer lediglich kleine "inkrementale" Änderungen, jedoch keine mit unüberwindbaren Unsicherheiten verbundenen revolutionäre Änderungen im Hinblick auf vergangene, problemverwandte Entscheidungen.¹⁴

Wie Miller in den 50er Jahren anhand experimenteller Untersuchungen nachwies, beschränkt sich zudem die spontane Kapazität menschlicher Informationsverarbeitung im Kurzzeitgedächtnis auf sieben (plus/minus zwei) Einzelkategorien.¹⁵ Eine Konfrontation mit einer über diese "magische" Grenze hinausgehende Anzahl an informatorischen Entscheidungsvariablen erzeugt durch eine Überlastung des Wahrnehmungsvermögens eine Form kognitiven Stresses beim Entscheider. Dieser Wahrnehmungsüberlastung wird in der Fachliteratur als "information overload" bezeichnet,

Wirtschaftssubjekte haben demnach zwar den Willen rational zu handeln, sie besitzen jedoch nicht ausreichend Informationen bzw. nicht ausreichend rationales Denkvermögen zu vollständiger Informationsverarbeitung. In dem Maße, in dem die rationale Informationsverarbeitung an die Grenzen menschlicher Denkleistung stößt, gewinnt das Wissen eines Menschen, verstanden als individuelle Lernerfahrung, als Mittel zur Komplexitätshandhabung bspw. in betrieblichen Entscheidungen vor dem Hintergrund des Informationsüberflusses an Bedeutung.

⁹ Vgl. Simon (1957), S. 81

¹⁰ Vgl. Kirsch (1998), S. 181: Mit dem Attribut "synoptisch" wird unterstellt, dass der Entscheider über einen uneingeschränkten Überblick über alle Alternativen, alle möglichen Umweltsituationen und alle möglichen Ergebnisse der Alternativen verfügt.

¹¹ Vgl. Braybrooke/Lindblom (1963), S. 113

¹² Vgl. Picot/Reichwald/Wiegand (1998), S. 83: Der radikale Konstruktivismus beruht auf der neurobiologischen Erkenntnis, dass jede Form von Verstehen nicht ein getreues Abbild der Wirklichkeit, sondern durch die kognitiven Prozesse des Nervensystems lediglich ein internes Konstrukt der Außenwirklichkeit darstellt, das kognitive System des Gehirns demnach also eine operative Geschlossenheit aufweist.

¹³ Vgl. Kirsch (1998), S. 183 f.

¹⁴ Vgl. Kirsch (1998), S. 185

¹⁵ Vgl. Miller (1956)

2.1.3 Digitale Informations- und Kommunikationstechnik

Den zentralen Motor des Wandels von der Industrie- zur Informationsgesellschaft bilden die Entwicklungen im Rahmen der Informations- und Kommunikationstechnik (IuK-Technik). Dieser Fortschritt kann an drei Eckpunkten festgemacht werden: erstens, am Kapazitätenzuwachs der elektronischen Daten- und Informationsverarbeitung; zweitens, an der Ausweitung der elektronischen Vernetzung; drittens am Einzug der Digitalisierung in den Bereich der Signalübertragung.¹⁶

Eine Begleiterscheinung des durch die Verbreitungen digitaler IuK-Techniken ausgelöste "Wandel zum digitalen Technologieparadigma"¹⁷ ist die Ausprägung neuer Funktionen durch die vielschichtige Integration, Anreicherung und Ausdifferenzierung der Technologien in verschiedenen Anwendungsbe- reichen. Ein spezielles Phänomen im Rahmen der Digitalisierung bilden dabei die sowohl technologi- schen, wertschöpfungs- und nachfragebezogenen Verschmelzungseffekte im Rahmen der technologieaffinen Märkte für Telekommunikations-, Medien- und Informationstechnologieprodukte. Diese Verschmelzungseffekte werden häufig unter dem Begriff der "digitalen Konvergenz" subsu- miert.¹⁸

Ergebnis dieser Verschmelzungseffekte ist zum einen eine strukturelle Verbindung der bislang ge- trennten Märkte in den genannten Branchen.¹⁹ Ein Beispiel für eine solche Branchenverschmelzung bildet das Angebot für Anwendungen und Dienstleistungen im digitalen Mobilfunk, mit einem Spekt- rum von klassischen Telekommunikationsaufgaben wie Netzzugang und -betrieb, über inhaltsorien- tierte Medienprodukte wie mobile Nachrichtenportale bis hin zu Transaktionsdienstleistungen wie virtuelle mobile Marktplätze und Auktionen. Zum anderen ermöglicht die die technologische Ver- schmelzung von Digitalisierung, Vernetzung und Informationsverarbeitung die Herausbildung neuer Typen und Eigenschaften von Übertragungsmedien und -technologien.²⁰ Das Internet verbindet bspw. typische Eigenschaften klassischer Medien zur Individualkommunikation, wie im Falle von email- und Chatanwendungen, mit Eigenschaften publizistischer Massenkommunikation, bspw. in Form von elektronischen Büchern, Zeitungen oder Nachrichtensendungen.

Neben den Unterscheidungsmerkmalen der Multimedialität, Aktualität sowie ihrer zeitstabilen Verfüg- barkeit gegenüber traditionellen Medienformaten wie Druckerzeugnisse und Rundfunk, zeichnen sich digitale IuK-Techniken aufgrund ihrer dualen Kommunikationsfunktion durch ein hohes Maß an Inter- aktivität zwischen Distributor und Rezeptor aus.²¹ Ein Kunde kann nicht nur individuell adressiert werden, sondern besitzt ausserdem bspw. die Möglichkeit bequem und einfach Bewertungen über ein

¹⁶ Vgl. Zerdick et al. (2001), S. 150

¹⁷ Thielmann (2000), S. 10; in Anlehnung an Colombo/DangNguyen/Perrucci (1997) charakterisiert Thielmann ein technologisches Paradigma als "Modell einer Reihe technologischer Probleme, das auf ausgewählten naturwis- senschaftlichen Prinzipien und Basistechnologien beruht und als Definition relevanter Fragen ein Untersuchungs- bzw. Lösungsmuster bietet." Das "digitale Technologieparadigma" beschreibt darauf aufbauend den Leitcharakter der Digitalisierung von Informationen für die Innovationsentwicklung und die Marktdurchdringung der Informati- ons- und Kommunikationstechnik in zahlreichen Marktbereichen.

¹⁸ Vgl. u.a. EU (1999); Thielmann (2000) S. 9; Keuper (2002), S. 605 ff.

¹⁹ Vgl. Zerdick et al (2001), S. 141

²⁰ Vgl. Keuper (2002), S. 611 ff.; Hess (2002), S. 572 ff.

²¹ Vgl. Hess (2002), S. 577 f.

Produkt abzugeben, sich mit anderen Kunden auszutauschen oder spezifische Serviceleistungen in Anspruch zu nehmen. Die vielschichtigen Kommunikationsvorgänge über solche Rückkopplungsmechanismen können wiederum aus Sicht eines Produkt- bzw. Dienstbieters strukturiert erfasst und für eine analytische Bewertung des Kunden sowohl auf aggregierter Marktebene als auch auf Ebene der individuellen Person des Kunden verwendet werden.²²

2.2 Individuell orientierte betriebliche Handlungsmöglichkeiten

Um der veränderten Rolle des Individuums in der Gesellschaft Rechnung zu tragen, bzw. diese für das Unternehmen erfolgswirksam auszunutzen, eröffnen sich aus kommerzieller Sicht grundsätzlich zwei Handlungsfelder: zum einen ein an den individuellen Lernerfahrungen ausgerichtetes Management des Mitarbeiters als Wissensressource des Unternehmens (⑤), zum anderen eine umfassende Käuferorientierung durch ein individuell differenziertes Produktangebot im Rahmen der Wettbewerbsstrategie (④). Eine spezielle Variante des differenzierten Produktangebots ergibt auf Märkten für weitgehend vollständig digitalisierbare Produkte. Wie nachstehend eingehend untersucht wird, zerfällt auf solchen Märkten nicht nur das einzelvermarktungsfähige Dienstleistungsangebot von traditionell auf diesen Märkten aktiven Unternehmen, sondern es entstehen u.a. auch neue Formen kundenindividueller Erlösconfigurationen und damit Neuausrichtungen des betrieblichen Geschäftsmodells (⑥).

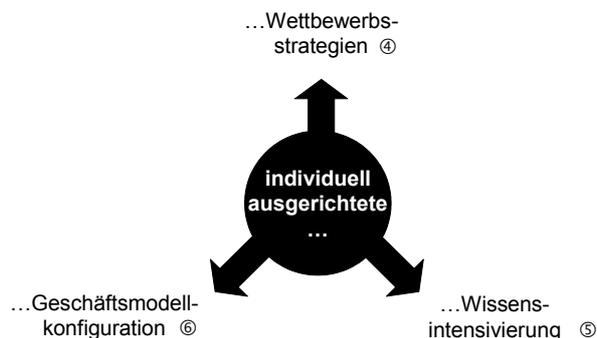


Abbildung 3: Individuelle Strategieorientierungen

2.2.1 Wettbewerbsstrategien und Individualität

Ebenso wie der wirtschaftliche Aufbau der westlichen Welt nach dem zweiten Weltkrieg, wie im ersten Abschnitt dargestellt, wohlstandsbedingte Auswirkungen auf die Gesellschaftsstruktur hatte, haben sich wachstumsbedingt auch die angebotsseitigen Strukturen des ökonomischen Wettbewerbs verändert. Auf den Angebotsmärkten nach dem zweiten Weltkrieg standen zunächst Aspekte der (vorwiegend kurzfristig orientierten) Rationalisierung und Organisation der Produktion im Vordergrund der Überlegungen der Überlebenssicherung.²³ Die zentrale Planungsaufgabe des Managements lag in der

²² Derartige Möglichkeiten zur strukturierten Erfassung von Kundenbedürfnissen und -merkmalen auf der Grundlage von Verhaltensbeobachtungen oder Befragungen sind Gegenstand von Kapitel 3.

²³ Vgl. Haertsch (2000), S. 49

Budgetierung der industriellen Produktion und der Kontrolle der Produktivität nach tayloristischen Prinzipien weitgehend variantenfreier Massenproduktion.²⁴ Als Synonym für die Serien- und Fließbandfertigung zu Beginn des Jahrhunderts gilt das erstmals 1908 vorgestellte "T-Modell" von Ford, von dem in danach 19 Jahren insgesamt über 15 Mio. Exemplare in nahezu unveränderter Baureihe produziert wurden.²⁵

Die zunehmende Diversifikation der Konsumentenbedarfe, der daraus resultierende Anstieg der Kapitalintensität langfristiger Investitionen und die aufkeimende Internationalisierung der Wettbewerbsstrukturen lenkte ab den 60er Jahren den Fokus auf längerfristige Prognosemethoden auf der Grundlage gegenwarts- und vergangenheitsbezogene Trendextrapolationen.²⁶ Globalwirtschaftlichen Diskontinuitäten – ein zentrales, prägendes Ereignis der Dekade war die globale Ölkrise 1973 – aber auch zunehmende Machtverlagerung in Richtung der Käuferschichten, machten ab den 70er Jahren einen grundlegenden Perspektivenwandel in der betrieblichen Planung notwendig. Die zunehmende Unsicherheit der kurzfristigen Zukunft konnte nicht mehr durch eine vergangenheitsorientierte Strukturfortschreibung bewältigt werden, sondern verlangte eine systematische Erforschung zukünftiger und damit strategischer Handlungsmöglichkeiten auf der Grundlage einer "systematischen Analyse der zukünftigen Chancen und Risiken in der Unternehmensumwelt".²⁷ Vor diesem Hintergrund identifizierte Porter 1980 zwei grundlegende Wettbewerbsstrategien, mittels derer ein Unternehmen Vorteile gegenüber Konkurrenten im Branchenwettbewerb erzielen kann: die Optimierung der betrieblichen Kostenstrukturen zur Erzielung eines Preisvorteils gegenüber Konkurrenten; die bedürfnisgerechte Bedienung von Kundengruppen durch ein differenziertes Produktangebot. Nach klassischem Verständnis der industriellen Fertigung muss zwischen den beiden Strategiealternativen eine strenge Auswahlentscheidung getroffen werden. Hintergrund bilden der zwangsläufige Anstieg der Komplexität in der Produktion bei Erhöhung der Variantenvielfalt und dadurch der Anstieg der direkten Kosten der Produktion, der Anstieg der Transaktionskosten aufgrund von Variantenplanung und Abstimmung sowie die Reduktion von Verbund- und Skaleneffekte.

Im Zuge der Effizienzpotenziale in Bezug auf den betrieblichen Informationsfluss durch die dargestellten Entwicklungen in der Informations- und Kommunikationstechnologie muss allerdings diese Alternativenauswahl in dreifacher Hinsicht neu bewertet werden:

1. Durch das breite Informationsangebot und dessen einfache und schnelle Zugänglichkeit im Rahmen digitaler Medien erhöht sich potenziell die Angebotstransparenz beim Kunden. Dadurch verringert sich der Preissetzungspielraum des Anbieters für, gemessen am Konkurrenten, homogene Produkte. Die Folge ist, dass sinnbildlich der Platz an der Spitze der Kostenführerschaft schrumpft und ein zusätzlicher Differenzierungsdruck für alle Wettbewerber der Branche entsteht.²⁸

²⁴ Vgl. Bea/Haas (1995), S. 11 sowie Hinterhuber (1992), S. 23

²⁵ Charakteristisch für das Planungsprinzip ist die Aussage von John Ford hinsichtlich einer Variation der Baureihe: «Sie bekommen jede Farbe - Hauptsache, sie ist schwarz!»

²⁶ Vgl. Bea/Haas (1995), S. 12

²⁷ Haertsch (2000), S. 49

²⁸ Vgl. Porter (2001), S. 7 ff.

2. Digitale IuK-Techniken verbessern nicht nur die Informationstransparenz im Rahmen der Produktvermarktung, sondern erhöhen auch potenziell die Organisationseffizienz im Rahmen betrieblicher Leistungsprozesse. Bei gleichem Kostenniveau kann dadurch ein höheres Maß an Prozesskomplexität bewältigt werden.²⁹
3. Wie bereits in Abschnitt 2.1.3 angesprochen, erhöhen digitale IuK-Techniken die Interaktivität zwischen Kunden und Anbietern. Dadurch erhöht sich potenziell auch die Transparenz der individuellen Bedürfnisse eines Kunden gegenüber dem Anbieter.

Ergebnis ist, dass als Folge der Informationswirtschaft die klassische Massenproduktion durch die Fertigung kundenindividueller Güter abgelöst wird, die zu einem zu einem Kostenniveau vergleichbarer Massengüter angeboten werden können.³⁰ Diese Fertigungsform wird als "Mass Customization" bezeichnet, als eine hybride Wettbewerbsstrategie aus Massenproduktion ("Mass Production") und Einzelfertigung ("Customization").

2.2.2 Wissensintensivierung und Individualität

Die Entwicklungen und kapazitären Erweiterungen in den Informations- und Kommunikationstechnologien ermöglichen einen Abschied vom traditionellen Kompromiss zwischen Reichhaltigkeit und Reichweite in Information und Kommunikation.³¹ Durch die Möglichkeiten von Breitbandübertragung, multimedialer Informationsaufbereitung und digitaler Interaktivität können mehr Informationen schneller und besser übertragen werden. Diese Verbesserung der Interaktionsmöglichkeiten führt insgesamt zu einer Steigerung des betrieblichen informations- und Kommunikationsaufkommens. Weltweit wird etwa mit einer Verdopplung des auf digitalen Datenträgern gespeicherten Datenvolumens alle 20 Monate gerechnet.

Durch die Zunahme der verfügbaren Informationen vergrößert sich gleichzeitig das aufgrund der angesprochenen Beschränkungen menschlicher Rationalität und Verarbeitungskapazität bestehende Auswahlproblem entscheidungsrelevanter Prämissen. Somit verliert im Zusammenhang mit den wachsenden Ansprüchen an technologischen Grundfertigkeiten, der Voraussetzung kreativen und komplexen Denkens für die Bewältigung der Dynamik in Wettbewerb und Branchenumfeld die menschliche Arbeitskraft den Charakter einer "Commodity".³² Forschungsintensive Produkte, komplexe organisatorische Abstimmungsprozesse und dynamische Rahmenbedingungen setzen heute die Maßstäbe der wirtschaftlichen Entwicklung in einer von Wissen geprägten Gesellschaft.³³

Wissen stützt sich also auf Daten und Informationen, ist jedoch im Vergleich breiter, tiefer und reicher.³⁴ Es entsteht durch eine personenabhängige Interpretation und Reproduktion der individuell

²⁹ Vgl. Piller (2001), S. 107 f.

³⁰ Piller (2001), S. 200

³¹ Vgl. Evans; Wursterer (2000), S. 37

³² Haertsch (2000), S. 14

³³ Vgl. Drucker (1994); Kübel (1990), S. 24

³⁴ Vgl. Davenport/Prusak (2000), S. 5

wahrgenommenen Problemsituation durch einen Entscheidungsträger auf der Grundlage seiner erlernten und erfahrungsgeprägten Denkkategorien und Erklärungsmuster.³⁵ Einen wesentlichen Faktor für diese interpretatorische Problemkategorisierung spielen die individuellen, sprachlichen Begriffsformen, auf deren Basis Individuen die Realität beschreiben und kommunizieren.³⁶ Solche sprachlichen Regelkonstrukte rekurrieren nach Wittgenstein im Wesentlichen auf die heterogenen "Sprachspiele" – gemeint sind sprachliche Bedeutungskonventionen, geprägt aus dem sozialen, sowohl privaten als auch beruflichen Lebensumfeld – mit denen sich ein Individuum im Laufe seines Lebens konfrontiert sieht.³⁷ Sprachlichen Konventionen repräsentieren eine zentrale Filterfunktion der Problemwahrnehmung und -definition für einen Entscheidungsträger und bilden damit den Kontext für die problemgebundene Entwicklung und Anwendung von Wissen.

Zwar können auch Organisationen gegenüber den originären Lebenswelten ihrer Mitglieder eigene Muster solcher sprachlicher Bedeutungskonventionen ausbilden.³⁸ Jedoch entstehen nur in dem Ausmaß, in dem eine Unternehmensführung durch das Schaffen spezifischer Sprach- und Ausdrucksformen die Kontexte ihrer Mitarbeiter zu prägen in der Lage ist, in den unternehmerischen Grenzen kollektive Wissens Elemente. Insofern repräsentieren die individuellen Persönlichkeiten der Mitglieder einer Organisation den zentralen Produktivfaktor für die Handhabung betrieblicher Probleme.³⁹

2.2.3 Geschäftsmodelle und Individualität

Die wachsende Bedeutung digitaler Medien hat nicht nur eine zunehmende Wettbewerbsbeziehung zwischen traditionellen und digitalen Medienmärkten zur Folge, sondern impliziert außerdem eine zunehmende Substituierbarkeit des Trägermediums. Diese Substitutionstendenz kann in zweifacher Hinsicht interpretiert werden: Aus Nachfragesicht lässt sich im Sinne der oben erörterten Indifferenz der Mediennutzung ein Trend in Richtung audiovisueller Informationsrezeption nachzeichnen. Aus Sicht der Inhaltsproduktion kann die Desintegration von Inhalten und Trägermedium als Resultat einer digitalen und vermehrt medienneutralen Inhaltserzeugung zurückgeführt werden.

In dem Maße, in dem sich Informationsgüter unabhängig von einem spezifischen Trägermedium erzeugen, bündeln und distribuieren lassen, zerfällt das vermarktungsfähige Angebot eines Anbieters solcher (traditionell kombiniert vermarkteter) Güter in die einzelvermarktbareren Leistungen der Inhaltserzeugung, der Bündelung und der kontextspezifische Kombination oder Aggregation von Inhalten sowie des Betriebs und des Zugangs zu digitalen Kommunikationsnetzen als Trägermedien im Inhaltsvertrieb.

³⁵ Vgl. Kirsch (1998), S. 183 f.

³⁶ Vgl. Staab (2002)

³⁷ Vgl. Wittgenstein (1984)

³⁸ Vgl. Habermas (1998)

³⁹ Vgl. Kübel (1990), S. 24

Das inhaltliche Überangebot sowie die traditionelle Gratismentalität des Informationsangebots im Internet erlauben wenn überhaupt nur geringe direkte Erlöse im Inhaltsvertrieb.⁴⁰ Dennoch übernehmen digitale Geschäftsbereiche – angesichts bereits spürbarer Veränderungen der Rezeptionsgewohnheiten der Mediennutzer in Richtung audio-visuell aufbereiteter Inhalte sowie zunehmender Kannibalisierungseffekte zwischen Onlinemedien und traditionellen Mediensegmenten – eine strategische Schlüsselbedeutung für das langfristige Überleben der Medienbranche.⁴¹ Aktuelle Untersuchungen zeigen, dass Rezipienten auch in Onlinemedien grundsätzlich bereit sind, für das Informationsangebot zu bezahlen, jedoch nur dann, wenn damit für sie hinsichtlich ihres Informationsstands bzw. der Bequemlichkeit der Informationsselektion und -rezeption gegenüber alternativen Medienprodukten sowohl in traditionellen als auch in digitalen Vertriebsmärkten ein Zusatznutzen verbunden ist.⁴² Ein solcher Rezeptionsvorteil kann vor dem Hintergrund des bereits beschriebenen Überangebots an Informationen insbesondere durch eine automatische Konfiguration des Dienstangebots in Bezug auf die Nachfragesituation eines Nutzers erreicht werden. Digitale Medien sind dabei in der Lage, nicht nur die Zeit und das Verhalten, sondern auch individuelle Präferenzen und den aktuellen Aufenthaltsort eines Nutzers als Grundlage einer solchen Konfiguration verwenden.

2.3 Zwischenfazit: Individualität und Individualisierung

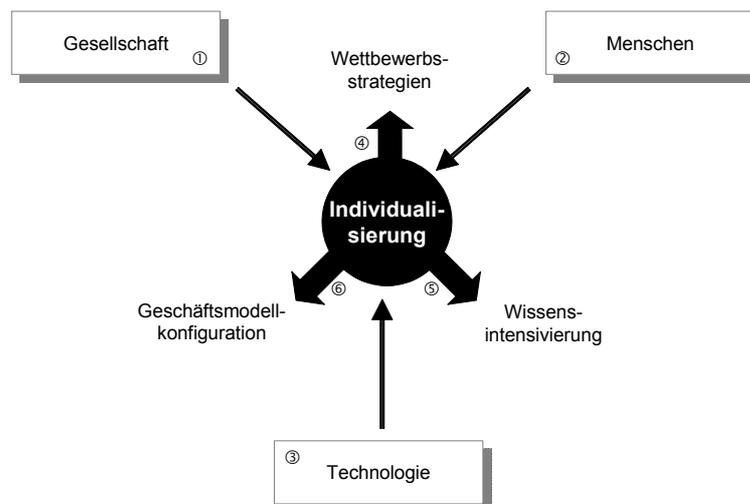


Abbildung 4: Synthese - Rahmen- und Gestaltungsbedingungen der Individualisierung

Als Zwischenergebnis der mehrdimensionalen Einführungsbetrachtung (vgl. Abbildung 4), lässt sich feststellen, dass sich die Individualität eines Menschen sowohl gesellschaftlich als auch ökonomisch zu einer Form "kollektiver Ideologie" entwickelt hat, die einerseits aus der Komplexität des postindustriellen Alltags heraus geboren ist, andererseits aber auch wohlstandsbedingtes Sinnbild eines modernen selbstbewussten Lebensstils darstellt. Im Weiteren sollen Handlungsmöglichkeiten untersucht

⁴⁰ Vgl. Mings/White (2000), S.64 f.

⁴¹ Vgl. u.a. Media Perspektive (2002), S. 64

⁴² Vgl. u.a. Hass (2002), S. 137, ECIN (2002) und ECIN (2003)

werden, um diesen Trend zur Individualisierung kommerziell in Form individualisierter Kunden-Absatz-Beziehungen ausnutzen zu können.

3 Individualität und individuelle Identifikation

Die traditionelle Marktforschung zielt darauf ab, über Umfragen und Beobachtungen aus der breiten Schicht der potenziellen Käuferschaft statistische Aussagen über die Struktur relevanter Kundenbedürfnisse zu ermitteln.⁴³ Die Interaktivität digitaler IuK-Technik eröffnet dagegen die Möglichkeit, Kundenbedürfnisse auf individueller Ebene in Form von Kundenprofilen zu identifizieren, die in ihrer Präzision weit über statistische Methoden hinausgehen.⁴⁴ Entsprechend der Qualität der über den Nutzer ermittelten Daten kann dabei unterschieden werden zwischen bspw. über überprüfte demographische oder abrechnungsbezogene Daten eindeutig identifizierenden Profilen, pseudonymen Profilen bspw. durch eine interessenbezogene Registrierung und anonymen Profilen auf der Grundlage verhaltensbezogener Nutzungsdaten (vgl. Abbildung 3).

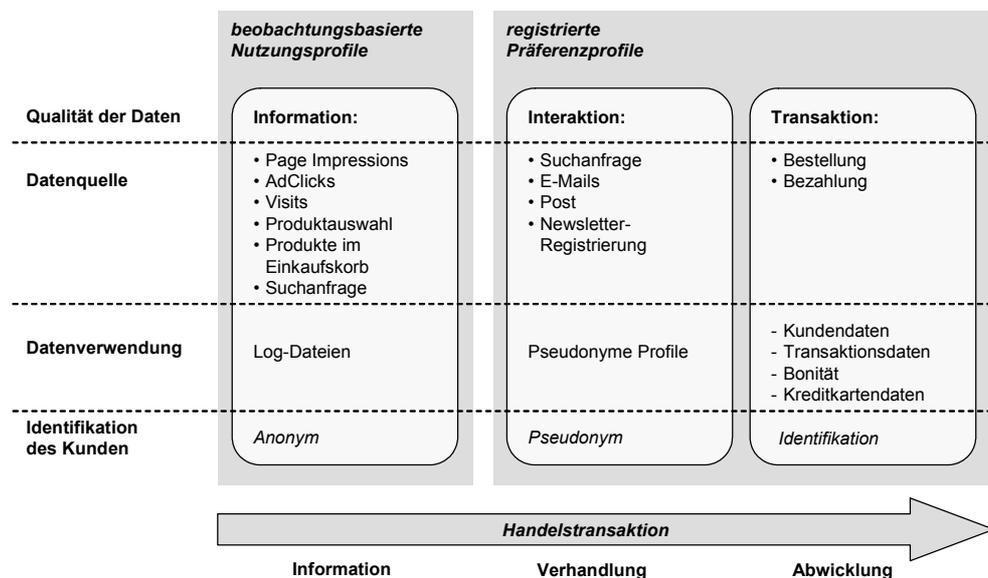


Abbildung 5: Qualität von Benutzerprofilen⁴⁵

Im Folgenden werden die Möglichkeiten im Rahmen der freiwillig – entweder eindeutig oder anonym – vom Kunden vorgenommenen Registrierung von Präferenzprofilen (Kapitel 3.1) sowie im Rahmen der verhaltensbezogener Ableitung von Nutzungsprofilen (Kapitel 3.2) vorgestellt. Insofern Kundeninteressen in Bezug auf Profilinformationen denen eines Online-Anbieters diametral gegenüberstehen,⁴⁶

⁴³ Vgl. Meffert (2000), S. 93 und 147 f.

⁴⁴ Vgl. Merz (2000), S. 234

⁴⁵ In Anlehnung an Merz (2000), S. 235

⁴⁶ Für Onlineanbieter stellen Profilinformationen zwar wertvolle Markt- und Kundeninformationen dar. Deren automatische Erhebung birgt jedoch die Gefahr eines Eingriffs in die geschützten Persönlichkeitsrechte des Kunden.

wird abschließend auf datenschutzrechtliche Bestimmungen im Rahmen der Kundenprofilermittlung eingegangen (Kapitel 3.3).

3.1 Registrierung und Präferenzprofile

Die Verwendung von Personendaten zur Individualisierung des Produktangebots setzt zweierlei voraus: erstens, die Ermittlung der Profildaten der jeweiligen Person; zweitens, die Speicherung der Profildaten und die Wiedererkennung der jeweiligen Person. Im Folgenden werden zunächst kurz die Mechanismen der Ermittlung von kundenindividuellen Profilen auf der Basis freiwilliger Angaben des Nutzers und die damit verbundenen Probleme erläutert. Im zweiten Abschnitt wird auf Mechanismen zur Wiedererkennung eines Kunden bei dessen mehrmaligem Aufruf eines Internetdienstes eingegangen.

3.1.1 Erhebung von Kundenprofilen durch Befragung

Die Interaktivität digitaler Kommunikationsmedien als Distributionskanal für digitale und Informationsgüter vereinfacht aus Sicht der empirischen Marktforschung die Möglichkeit der Kundenkommunikation sowohl in Bezug auf die technische Abwicklung als auch in Bezug auf die Reichweite der Erhebung von Kundeninformationen im Rahmen von Markt- und Präferenzanalysen.⁴⁷ Dadurch erschließt sich die Möglichkeit nicht nur zur Erhebung stichprobenhafter Marktdaten für eine statistische Aggregation der Marktstruktur, sondern darüber hinaus für ein umfassendes Kundenmanagements in Form der Speicherung von Einzelkundenprofilen auf der Grundlage von mediatisiert in digitaler Form erhobenen demographischen oder präferenzbezogenen Angaben. Eine im Internet mittlerweile geläufige Form der Erhebung von individuellen Kundenprofilen bildet eine obligatorische, häufige kostenfreie Registrierung eines individuellen Nutzerzugangs für die Nutzung von Informationsangeboten.

Bereits Shapiro und Varian erkannten allerdings das Dilemma im Rahmen der Registrierung von Personen- und Präferenzprofilen: Mangels Kontrollierbarkeit durch den Anbieter ist die Qualität freiwilliger Angaben von Kunden in der Regel sehr schlecht und in hohem Grade abhängig vom kalkulierten Nutzen des jeweiligen Kunden aus der Preisgabe seiner Daten.⁴⁸ Eine Untersuchung des Internetnutzungsverhaltens an der amerikanischen Vanderbilt University deckte 1999 auf, dass 94% der beobachteten Internetnutzer generell keine persönlichen Informationen in einer Internetseite hinterließen, und 40% der registrierten Informationen falsche Angaben enthielten. Wie in Abbildung drei deutlich wird, kann die Qualität und Verbindlichkeit von Profildaten jedoch durch die Schaffung dauerhafter Transaktionsbeziehungen gesteigert werden: Während Informationsdienste lediglich eine anonyme Identifikation bspw. auf Grundlage von Logfile-Analysen erlauben, setzen interaktive Dienste wie Suchanfragen, Mailedienste oder Newsletter zumindest ein pseudonymes Profil, Transaktions-

⁴⁷ Vgl. Bensberg (2001), S. 2

⁴⁸ Shapiro/Varian (1999), S. 35

dienste wie Shopping-Portale schließlich eine eindeutige und nachgewiesene Identifikation der Person voraus.

Der Online Buchhändler amazon.de verwendet beispielsweise ein zweistufiges Registrierungsverfahren: Zur Teilnahme an den verschiedenen Formen der Produktempfehlung kann sich ein Nutzer zunächst ein Profil seiner Präferenzen anlegen. Dazu muss eine gültige e-Mailadresse angegeben sowie eine zunächst nicht weiter authentifizierte Namensangabe gemacht werden. Die Präferenzen werden dann in einer vierseitigen Auswahlliste von den bevorzugten Produktbereichen auf themen- und schließlich produktspezifischer Ebene eingeschränkt (vgl. Abbildung 4). Zum Abschluss einer Transaktion müssen im zweiten Schritt dann eine gültige Rechnungs- und Lieferanschrift bzw. eine Einzugsermächtigung für ein überprüftes Kreditkarten- oder Bankkonto hinterlegt werden.

Abbildung 6: Profilregistrierung bei Amazon

Neben dem angesprochenen Problem der mangelnden Überprüfbarkeit der Validität von im Rahmen mediatisierter, reaktiver Befragungen erhobenen Kunden- und Präferenzinformationen begründen sich weitere Probleme in Bezug auf die Datenqualität in der Natur der Befragung als Erhebungsform. Der Vorteil einer Individualisierung der Kundenbeziehung im Sinne eines "one-to-one"-Marketingansatzes besteht darin, aus der Mengen potenzieller Kunden diejenigen zu identifizieren, die eine starke Produkt- oder Dienstleistungszufriedenheit und damit eine hohe Wiederkaufswahrscheinlichkeit aufweisen.⁴⁹ Durch die freie Zugänglichkeit einer mediatisierten Befragung im Falle einer freiwilligen Dienstregistrierung können jedoch zum einen Selbstselektionseffekte entstehen, die bspw. aufgrund der notwendigen technischen Bedienkompetenz des Mediums für die Teilnahme an der Befragung potenziell produktaffine Kunden aus dem Teilnehmerkreis ausschließen und somit die Aussagekraft der Daten des registrierten Kundenstamms verfälschen.⁵⁰ Zum anderen sind die per einmaliger Befragung und Registrierung ermittelten und gespeicherten Profile im Zeitablauf statisch.⁵¹ Lebenslaufbedingte Änderungen der Präferenzen, bspw. bei einem Wechsel des Arbeitsplatzes oder Veränderungen in der familiären Situation eines Kunden, bzw. Veränderungen des verfügbaren Ange-

⁴⁹ Vgl. Peppers/Rogers (1996), S. 32

⁵⁰ Vgl. Bensberg (2001), S. 35

⁵¹ Vgl. Mobasher (2002), S. 2

bots, bspw. durch Produktinnovationen mit bislang unbekanntem Ausprägungen der Produktmerkmale, können daher nicht ausreichend berücksichtigt werden.

3.1.2 Identifikation des Kunden über registrierte Profile

Um eine dauerhafte Beziehung mit dem Kunden vor dem Hintergrund der Erhöhung der Produkttreue zu gewährleisten, muss die eindeutige Wiedererkennung des Kunden anhand des gespeicherten Kundenprofils gewährleistet werden. Die Identifizierung zwischen zwei Kommunikationspartnern im Internet erfolgt grundsätzlich über den 32-Bit Zahlencode der IP-Internet-Adresse. Während eines Kommunikationsvorgangs muss jedem Sender und Empfänger eine eindeutige Adresse zugeordnet sein. Speziell privaten Anwendern, die über einen Dienstanbieter auf das Internet zugreifen, wird allerdings in der Regel für jede Sitzung eine dynamische IP-Adresse aus einem Adresspool des Zugangsdienstleisters zugeordnet, wodurch sie sitzungsübergreifend nicht mehr eindeutig allein anhand der IP-Adresse identifiziert werden können. Technisch bieten sich derzeit drei alternative Verfahrensweisen zur Nutzeridentifizierung an: die Verwendung clientseitiger Cookies, die Wiedererkennung des anbieterseitig gespeicherten Kundenprofils durch eine clientseitige Erweiterungssoftware des Browsers und die Registrierung von Profilinformatoren bei einer authentifizierten Drittinstanz.

1. Unter dem Begriff "Cookie" wird eine Erweiterung der HTTP-Spezifikation zum Austausch von Statusinformationen verstanden, standardisiert als "Request for Comment (RFC) 2109" im Rahmen der HTTP-Spezifikation.⁵² Cookies sind kleine Textdateien, die auf der Festplatte des Kunden gespeichert werden. Typischerweise enthält ein Cookie lediglich jeweils eine Identifikationsnummer zur Wiedererkennung eines Benutzers und einer Sitzung oder Transaktion.⁵³ Weitere standardisierte Attribute im Rahmen der Spezifikation RFC 2109 bilden der Verfallszeitpunkt eines Cookies, der Domainnamen des Webserver des Anbieters und ggf. der Pfadname der Dateireferenz durch die die Sendung des Cookie auslöst wird.⁵⁴ Ein Cookie wird durch den Webserver des Anbieters an den Browser des Nutzers, vor dem eigentlichen HTML-Dokument zur Speicherung übertragen und entweder bei Abfragen unterhalb des referenzierten Dateipfads oder alternativ, falls dieser nicht angegeben ist, bei jedem HTTP-Befehl des Browsers an den Server gesendet. Aus datenschutzrechtlichen Gründen legt die Cookie-Spezifikation jedoch fest, dass der Benutzer die Cookie-Funktionalität zur Wahrung seiner Anonymität jederzeit ausschalten kann. Somit bilden Cookies kein zuverlässiges Instrument für eine eindeutige Identifizierbarkeit eines Nutzers.
2. Zur Speicherung von Kundenprofilinformationen beim Dienstanbieter reichten Netscape, Firefly und VeriSign 1998 beim World Wide Web Consortium (W3C) die Spezifikation des "Open Profiling Standard" (OPS) als Träger der Profilinformatoren als Vorschlag für einen weltweiten Kundenprofilstandard ein. OPS legt Datenformate und Übertragungsverfahren fest, die die Online-Registrierung vereinfachen und das Problem der ungeschützten Ablage von Profilin-

⁵² Vgl. Kristol/ Montulli (1997)

⁵³ Vgl. Amberg/Wehrmann (2001), S. 15 f.

⁵⁴ Vgl. Merz (1999), S. 237

formationen lösen sollen.⁵⁵ Die Verwendung von OPS setzt eine zusätzliche Softwarekomponente im Internetbrowser voraus. Ein Kunde legt damit ein individuelles Profil an. Die Profilsoftware kontrolliert die vom Kunden einzeln festlegbare Freigabe jeder Profilinformativkategorie für jeden abgefragten Dienst.

- Die Speicherung von Profilinformativ bei einer Authentifizierungsinstanz soll zukünftig auf der Grundlage des "Platform for Privacy Preferences Protocol" (P3P) erfolgen, das im April 2002 als Spezifikation P3P 1.0 vom W3C veröffentlicht wurde.⁵⁶ Das P3P repräsentiert eine Protokollfamilie, die sich aus verschiedenen Teilstandards zur einheitlichen Aushandlung von Profildaten zusammensetzt.⁵⁷ Das P3P dient dabei sowohl zur Standardisierung in der Kommunikation zwischen Dienstleister und der authentifizierenden Drittinstantz, bspw. in Bezug auf Verhandlungsprotokoll und Vokabular für Profilinformativ und Datenschutzregeln, als auch hinsichtlich der Präferenzdefinition von Veröffentlichungsregeln zwischen Authentifizierungsinstanz und Kunde (vgl. Abbildung 7).

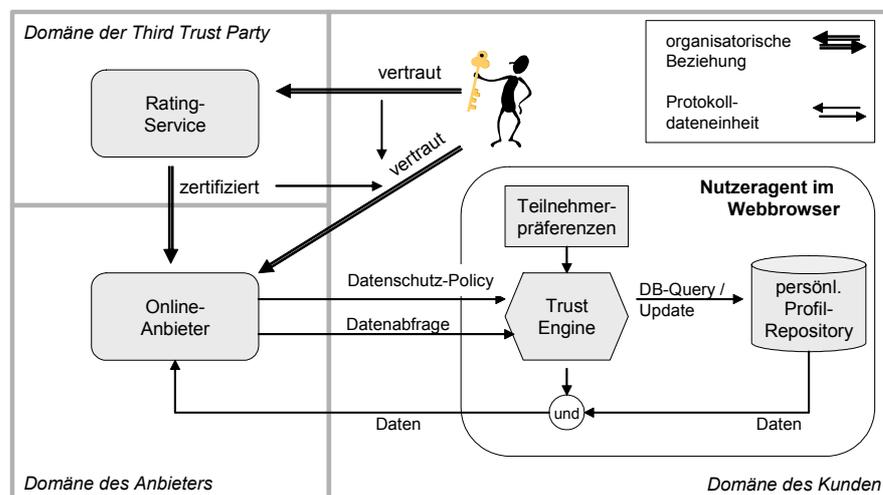


Abbildung 7: P3P Kommunikationsarchitektur⁵⁸

Praktische Umsetzungen, insbesondere in Bezug auf die Gestaltung der Authentifizierungsinstanz stehen allerdings noch aus.

3.1.3 Bewertung von freiwilligen Benutzerprofilen zur individuellen Kundenidentifikation

Demographische Angaben des Kunden zu Person, Alter, Wohnsitz, Beruf und ggf. Bankverbindung bilden grundsätzlich eine günstige Grundlage zur Identifikation eines Kunden. Allerdings kann auf diese Angaben ohne freiwillige Preisgabe des Kunden nicht anderweitig rückgeschlossen werden. Mangels Überprüfbarkeit sind die vom Kunden in der Praxis im Rahmen einer Profilregistrierung für Dienste im Internet gemachten Angaben in der Regel von nur geringer Aussagekraft für einerseits die Identifikation, andererseits auch – wie gezeigt – für die langfristigen Bedürfnisse eines Kunden. Ein

⁵⁵ Vgl. Merz (1999), S. 249

⁵⁶ Vgl. Cranor/Langheinrich/Marchiori/Presler-Marshall/Reagle (2002)

⁵⁷ Vgl. Cranor/Weitzner (2002)

⁵⁸ In Anlehnung an Merz (2000), S. 257

weiteres Problem in Verbindung mit der Verwendung freiwilliger Angaben des Kunden entsteht zudem im Zusammenhang mit dessen sitzungsübergreifender Wiedererkennung. Aufgrund seiner Deaktivierbarkeit im Browser des Nutzers bildet der Cookie Mechanismus kein verlässliches Verfahren zur Kundenidentifikation und findet deshalb heute in der Praxis nur noch in Ausnahmefällen Anwendung. Alternative Verfahren zur Wiedererkennung konnten sich jedoch demgegenüber bislang noch nicht durchsetzen. Standardisierungen wie die P3P-Familie stellen zwar ein ambitioniertes Vorhaben dar, stehen jedoch absehbar vor einigen Umsetzungshürden: P3P ist zum einen eine Standardisierung, die im Wesentlichen auf die regulatorischen Bedingungen der USA abzielt und daher einen sehr liberalen Ansatz darstellt. Im Hinblick auf die Geschäftspraktiken der USA, wo praktisch jede Form von Information frei gehandelt wird, stellt P3P dennoch eine regulatorische Einschränkung für Onlineanbieter und somit eine wesentliche Akzeptanzhürde dar. Demgegenüber ist es fragliche, ob sich P3P an den, in Europa im Vergleich zu den USA verordneten Mindeststandards für den Datenschutz anpassen lässt. Zum anderen unterliegt die reale Nutzung von P3P, wie in Abbildung 7 dargestellt, in hohem Maße dem Vertrauen der Nutzer in die tatsächlich praktizierte Einhaltung der offiziell bekundeten Datenschutzvereinbarungen durch Authentifizierungsinstanz und Anbieter. Mangels detaillierter Rechtsvorschriften in diesem Bereich ist es zukünftig vor allem Aufgabe der Authentifizierungsinstanz, zunächst ein solches Nutzervertrauen aufzubauen und durch Möglichkeiten der Einhaltungskontrolle bei Anbietern auch langfristig zu bestätigen.

Aufgrund der genannten Unzulänglichkeiten der individuellen Identifikation von Nutzern auf der Grundlage freiwilliger Angaben, wächst die Bedeutung von Verfahren zur Nutzeridentifikation auf der Grundlage der indirekten Beobachtung des Nutzerverhaltens. Insbesondere wird die Anwendung von Methoden des Data Mining auf Verhaltensdaten eine hohe methodische Qualität zur Ableitung von Nutzungsprofilen und auf deren Grundlage zur Ableitung von individuellen Kundenbedürfnissen beschieden. Im Folgenden werden diese Verfahren des Data Mining vorgestellt.

3.2 Web Usage Mining und Nutzungsprofile

Bei jedem Seitenaufruf im Internet hinterlässt ein Nutzer Spuren auf den Webservern der verschiedenen Anbieter. Jeder Webserver verfügt über eine Protokollkomponente, ein softwaretechnisches System, das sämtliche Serveraktivitäten registriert und chronologisch in meist unterschiedlichen anwendungsorientierten Protokolldateien, den so genannten "Server Logfiles", abspeichert. Neben dem sog. "Zugriffsprotokoll", das sämtliche HTTP-Elementaroperationen dokumentiert, werden im sog. "Browserprotokoll" die Typenbezeichnung des Browsers und des Betriebssystems des zugreifenden Clients, im sog. "Referentenprotokoll" die URL-Adresse der Ressource, die der Client in der vorhergehenden Interaktion angefordert hat und in einem Fehlerprotokoll ggf. Uhrzeit, Zieladresse des Verbindungsversuchs und Fehlerart aufgezeichnet.⁵⁹ Diese Protokolldateien können zum Gegenstand von systematischen Beobachtungen des Nutzerverhaltens werden, über die Rückschlüsse auf die Präferenzen von Nutzern gezogen werden können.

⁵⁹ Vgl. Schockert (2003); Bensberg (2001), S. 43 f.

Abbildung 8 zeigt den Zeilenausschnitt eines typischen Zugriffsprotokolls im Beispiel des Internetauftritts der Universität Göttingen.

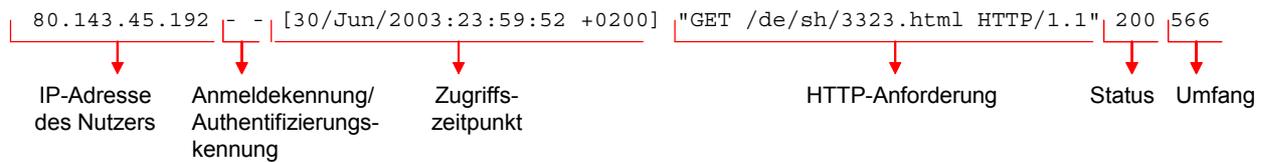


Abbildung 8: Struktur eines Server Logfiles

Im Zugriffsprotokoll werden folgende Angaben protokolliert: die IP-Adresse des anfragenden Nutzers; die ggf. übertragene Anmelde- und Authentifizierungskennung des Nutzers (üblicherweise leer); der Zugriffszeitpunkt der Serveranfrage, mit Angaben zu Datum (Tag, Monat, Jahr) und Uhrzeit (Stunde, Minute, Sekunde, Zeitzone); die HTTP-Anforderung des Zugriffs; Aufgezeichnet werden die Zugriffsmethode ("GET" für Downloads, "POST" für Uploads oder "HEAD" für Anfragen, ob sich die Datei geändert hat), der Zugriffspfad zur übertragenen Datei und das Serverprotokoll; der Status der Dateiübertragung ("200" für ordnungsgemäß abgeschlossene Übertragung OK, "304" für die Dateiübertragung bereits besuchter Seiten aus dem Browsercache des Nutzers); die Umfang der übertragenen Datei in Byte.

Durch die Vollständigkeit der Aufzeichnung sämtlicher durch Nutzer ausgelöste Aktivitäten bieten Server Logfiles eine geeignete Grundlage für die Analyse des Nutzerverhaltens im Sinne einer indirekten Beobachtung. Dabei sind grundsätzlich zwei verschiedene Analysemethoden denkbar: zum einen berichts-basierte Analysen, die auf der Grundlage von Häufigkeitskennzahlen aggregierte statistische Aussagen generieren; Gängige Protokollauswertungen beschäftigen sich unter anderem mit der Ableitung von Interaktivitätsmaßen, bspw. in Bezug auf die Anzahl bestimmter Bereiche im Informationsangebot einer Webseite oder die durchschnittliche Verweildauer eines Nutzers und der Ableitung von technischen Kennzahlen, bspw. in Bezug auf den Auslastungsgrad des Servers.⁶⁰ Zum anderen multivariate Methoden, speziell aus dem Bereich des Data-Mining, zur Überprüfung vorformulierter oder automatisch generierter Hypothesen.⁶¹ Für die Identifikation von individuellen Kundenpräferenzen spielen insbesondere letztgenannte Verfahren eine zunehmend bedeutsame Rolle zur Entwicklung und Überprüfung von Hypothesen in Bezug auf individuelle Verhaltensmuster im Rahmen der Gesamtheit der protokollierten Aktivitäten. Im Folgenden werden deshalb die Methoden des Data Mining in der Anwendung auf Server Protokolldateien vorgestellt.

3.2.1 Data Mining und Knowledge Discovery in Databases

Unter Data Mining ist ein Prozess zu verstehen, "der aus einer Datenmenge implizit vorhandene, aber bisher unentdeckte, nützliche Informationen extrahiert."⁶² Die Methodik des Data Mining rekuriert allgemein auf den Gestaltungsbedarf für Analysesysteme im Zuge der wachsenden Datenflut, zur

⁶⁰ Vgl. Meyer/Wingärtner/Jahke/Lieven (2001), S. 1

⁶¹ Vgl. u.a. Mobasher (2002); Spilino-poulou (2001); Meyer/Wingärtner/Jahke/Lieven (2001)

⁶² Vgl. Mertens (1994), S. 793

Untersuchung komplexer Zusammenhänge im Rahmen großvolumiger Datenbestände, die traditionelle statistische Verfahren nur noch unzureichend durchdringen. Im Bereich des Internets ergeben sich dabei zwei Einsatzfelder für das Data Mining (vgl. Abbildung 9):⁶³ zum einen, die strukturelle und inhaltliche Analyse von Internetseiten und -verweisen, das so genannte "Web Content Mining"; zum anderen die Analyse von Nutzeraktivitäten in Bezug auf die Seiteninhalte, das so genannte Web Usage Mining. Im Bereich des Web Usage Mining lassen sich wiederum zwei Anwendungsbereiche unterscheiden: die nicht personalisierte Nutzungsanalyse zur Identifikation von Verhaltensmuster bspw. zur Bildung von Rückschlüssen aus dem Nachfrageverhalten über die inhaltliche Qualität einer Webseite; die personalisierte Nutzungsanalyse zur Identifikation individueller Verhaltensmuster bspw. zur Erhöhung von Kundenzufriedenheit und Kundenbindung durch individuell orientierte Angebote.

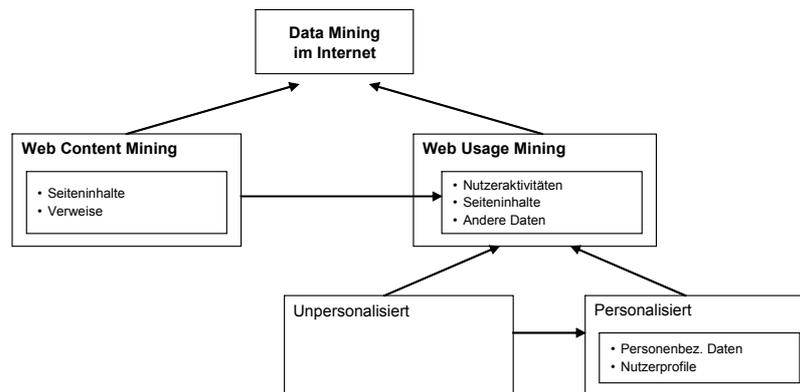


Abbildung 9: Anwendungsbereiche des Data Mining im Internet⁶⁴

Data Mining basiert im Wesentlichen auf multivariaten statistischen Verfahren zur Datenaufbereitung. Zum Instrumentarium gehören Klassifikationen bspw. auf der Grundlage von Diskriminanzanalysen oder induktiven Lernverfahren, lineare und nichtlineare Regressionen, Segmentierungen durch Gruppenbildung auf der Grundlage von statistischen Verfahren oder maschinellem Lernen sowie Abhängigkeits- und Abweichungsanalysen.⁶⁵ Data Mining dient zur Ableitung von Informationen aus einer Menge von Daten bspw. zum Zwecke der Ermittlung von Gestaltungsempfehlungen der Web-Site oder der Kontrolle der Seitenqualität.⁶⁶ Durch Anwendung von Methoden des Data Mining erfolgt aus Sicht der Semiotik eine Verdichtung von syntaktisch beschriebenen Daten, wie hier im Fall einer Webserver Protokolldatei, zu sachfeldorientierten Informationen und deren Anwendung vor dem Hintergrund des Gestaltungs- oder Kontrollziels also problembezogenes Wissen. Dieser Vorgang wird allgemein beschreiben als Prozess der Wissensidentifikation in Datenbanken (KDD).⁶⁷ Dieser Prozess untergliedert sich in mehrere Teilschritte (vgl. Abbildung 11): die Selektion und Transformation von Daten, die Ableitung von Datenmustern, deren Evaluation, Präsentation und Interpretation vor dem Hintergrund der zugrunde gelegten Hypothese.

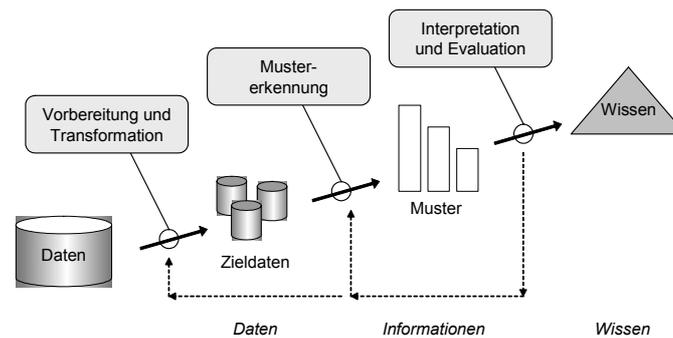
⁶³ Vgl. Spiliopoulou (2001), S. 491

⁶⁴ Vgl. Spiliopoulou (2001), S. 491

⁶⁵ Vgl. Bensberg (2001), S. 96

⁶⁶ Vgl. Meyer/Wingärtner/Jahke/Lieven (2001), S. 3

⁶⁷ Vgl. Frawley/Piatetsky-Shapiro/Matheus (2001); Fayyad/Piatetsky-Shapiro/Smyth (1996)

Abbildung 10: KDD⁶⁸

Bezugnehmend auf Analyse von Webserver Protokolldateien zur Identifikation individueller Verhaltensmuster müssen im Zuge eines solchen analytischen Prozesses der Wissensidentifikation im Rahmen der Vorbereitungsphase die protokollierten Daten zum einen von den für die Unterscheidung individuellen Verhaltens nicht relevanten Aktivitäten bereinigt werden; zum anderen müssen die Aktivitäten unterschiedlicher Nutzer getrennt und nach Aggregaten virtueller Kundenbesuche wie Sitzungen oder Transaktionen gruppiert werden.⁶⁹ Zur Mustererkennung im nächsten Schritt können dann auf Grundlage vorgegebener oder automatisch generierter Hypothesen über das Nutzerverhalten durch Data Mining Algorithmen signifikante individuelle Verhaltensmuster identifiziert werden. Die Ergebnisse werden im letzten Schritt, bspw. anhand von Kriterien wie Validität, Neuheit oder Verständlichkeit in Bezug auf ihren Wissensbeitrag zum sog. "Domänenwissen", also der allgemeinen Sachfeldkompetenz eines, in diesem Falle Webseitenbetreibers über seinen Zielmarkt, bewertet.⁷⁰

Insofern weniger die Frage einer ggf. automatisierten Ergebnismessung, als vielmehr die grundsätzliche Generierung solcher Ergebnisse in Bezug auf die Identifikation individueller Verhaltensmuster von Interesse ist, wird im Folgenden lediglich auf die Voraussetzungen und Möglichkeiten des Data Mining in den beiden erstgenannten Prozessteilschritten der Datenvorbereitung und -transformation sowie der Mustererkennung eingegangen.

3.2.2 Datenvorbereitung und -transformation von Webserver Protokolldateien

Im Rahmen der Datenvorbereitung können protokollierte Aktivitäten in einer Protokolldatei sowohl in zweifacher Hinsicht untersucht werden: Im Zuge einer vertikalen Datenselektion werden verschiedene Datensätze auf ihre Relevanz hin untersucht. Demgegenüber werden im Zuge einer horizontalen Selektion die Daten innerhalb jeweils einer protokollierten Elementaroperation des Webservers unter-

⁶⁸ In Anlehnung an Fayyad/Piatetsky-Shapiro/Smyth (1996), S. 41 und Bensberg (2001), S. 72: Fayyad et al. unterscheiden fünf Teilprozesse: Selektion, Vorbereitung, Transformation, Data Mining und Interpretation/Bewertung; Bensberg fasst die beiden Schritte Vorbereitung und Transformation zu einem Teilprozess zusammen, ergänzt jedoch die für eine vollständige Auswertung der Daten notwendigen Folgeschritte der Präsentation und Interpretation evaluierter Muster. Zur Vereinfachung werden hier im Weiteren die Teilschritte der Datenvorbereitung und -transformation nochmals zusammengefasst, ebenso wie die Schritte der Auswertung identifizierter Muster, also Evaluation, Präsentation und Interpretation.

⁶⁹ Vgl. Spiliopoulou (2001), S. 493

⁷⁰ Vgl. Fayyad/Piatetsky-Shapiro/Smyth (1996), S. 50; Bensberg (2001), S. 72

sucht.⁷¹ Der Webserver protokolliert generell sämtliche Aktivitäten, die eine Nutzeranfrage auslöst. Dazu gehören neben den tatsächlichen Seitenaufrufen auch bspw. die Ausführung von grafischen oder verdeckten funktionalen Seitenelementen, Seitenzugriffen von Suchmaschinen oder fehlerhafte Seitenübertragungen. Insofern solche Aktivitäten nicht bewusst von einem Nutzer aufgerufen werden, können sie im Rahmen der vertikalen Datenselektion zur Reduktion des Umfangs der zu analysierenden Daten herausgefiltert werden. Abbildung 12 zeigt am Beispiel eines Ausschnitts aus der Protokolldatei des Webserver für den Internetauftritt der Universität Göttingen drei Formen solcher filterbaren Aktivitäten: eine automatisch in Verbindung mit einer Seite aufgerufene Javascript Datei für funktionale Seitenbedienelemente (①); der automatische Aufruf einer eingebundenen Grafikdatei (②) sowie eine Transkriptionsregel des Webserver für die Umsetzung dynamischer Seitenadresse in der Gestalt statischer HTML-Links (③).

```

134.76.23.27 - - [30/Jun/2003:23:59:49 +0200] "GET /frames.js HTTP/1.1" 200 627 ← ①
134.76.23.27 - - [30/Jun/2003:23:59:50 +0200] "GET /images/leiste2.gif HTTP/1.1" 200 159 ← ②
134.76.164.115 - - [30/Jun/2003:23:59:50 +0200] "GET /show.php?lang=de&type=sh&id=13 HTTP/1.1" 200 552
212.253.2.201 - - [30/Jun/2003:23:59:50 +0200] "GET /de/sh/13.html HTTP/1.1" 200 552
134.76.164.115 - - [30/Jun/2003:23:59:50 +0200] "GET /show.php?lang=de&type=sh&id=13 HTTP/1.1" 200 552 ← ③
212.253.2.201 - - [30/Jun/2003:23:59:50 +0200] "GET /de/sh/13.html HTTP/1.1" 200 552
212.253.2.201 - - [30/Jun/2003:23:59:51 +0200] "GET /show_header.php?id=13&lang=de HTTP/1.1" 200 5346
217.0.71.184 - - [30/Jun/2003:23:59:52 +0200] "GET /show_kat.php?id=3013 HTTP/1.1" 200 27764
134.76.164.115 - - [30/Jun/2003:23:59:52 +0200] "GET /show.php?lang=de&type=sh&id=3323 HTTP/1.1" 200 566
80.143.45.192 - - [30/Jun/2003:23:59:52 +0200] "GET /de/sh/3323.html HTTP/1.1" 200 566
80.143.45.192 - - [30/Jun/2003:23:59:52 +0200] "GET /show_header.php?id=3323&lang=de HTTP/1.1" 200 5346
212.253.2.201 - - [30/Jun/2003:23:59:52 +0200] "GET /de/ue/13.html HTTP/1.1" 200 15662
134.76.164.115 - - [30/Jun/2003:23:59:52 +0200] "GET /show.php?lang=de&type=ue&id=13 HTTP/1.1" 200 15661
134.76.164.115 - - [30/Jun/2003:23:59:52 +0200] "GET /show.php?lang=de&type=ue&id=3323 HTTP/1.1" 200 16919
80.143.45.192 - - [30/Jun/2003:23:59:52 +0200] "GET /de/ue/3323.html HTTP/1.1" 200 16919
80.143.45.192 - - [30/Jun/2003:23:59:52 +0200] "GET /show_footer.php?id=3323&lang=de HTTP/1.1" 200 7265
80.143.45.192 - - [30/Jun/2003:23:59:52 +0200] "GET /images/schueler.gif HTTP/1.1" 200 4542 ← ②

```

Abbildung 11: Datenbereinigung im Zugriffsprotokoll

Nach dem Entfernen dieser Einträge müssen dann die übrig gebliebenen Elementaroperationen horizontal hinsichtlich der interessierenden Daten zerlegt werden, bspw. in Form der Selektion der Dateiadresse aus dem zusammengesetzten Attribut der HTTP-Anforderung.

Aufgrund der bereits angesprochenen Akzeptanzprobleme in Verbindung mit Browserregistrierung per Cookie oder Benutzerauthentifizierung per Registrierung repräsentieren die fehlenden Anmeldungs- und Authentifizierungskennungen eines Nutzers Fehlzeiten für die individuelle Identifikation eines Nutzers. Deshalb müssen für die Identifikation einzelner Nutzersitzungen oder Transaktionen sachliche oder zeitliche Zusammenhänge zwischen den protokollierten Elementaroperationen der Protokolldatei hergestellt werden.⁷² Eine Möglichkeit zur Transaktionsableitung auf der Grundlage zeitlicher Abgrenzung bietet sich bspw. durch Gruppenbildung in Form einer Clusteranalyse über die zeitliche Distanz zwischen Operationen, ausgelöst von derselben Internetadresse.⁷³ Als Proximitätsmaß zur Gruppenbildung wird vom Beobachter und Analysten dabei eine kritische Referenzdauer, bspw. auf der Basis der durchschnittlichen Sitzungsdauer einer beobachteten Testgruppe, festgelegt. Eine andere Form der Transaktionsableitung liegt in der rückwärtigen Pfadrekonstruktion zwischen verschiedenen Abfragen von derselben Zugriffsadresse bis hin zum Einstiegspunkt einer Webseite. Als

⁷¹ Vgl. Bensberg (2001), S. 134-137

⁷² Vgl. Bensberg (2001), S. 138

⁷³ Vgl. Bensberg (2001), S. 140

Analysegrundlage dienen, neben Internetadresse und Zugriffszeitpunkt aus dem Zugriffsprotokoll, die aufgezeichneten Navigationspfade im Referentenprotokoll.⁷⁴

Im Rahmen der Datentransformation werden die gruppierten Daten schließlich, bspw. in Form einer parametrisierten oder binären Vektorisierung, für die Anwendung der Algorithmen zur Mustererkennung aufbereitet. Dieser Teilschritt hängt maßgeblich von der gewählten Data Mining-Methodik ab, wobei bspw. Assoziations- und Sequenzanalysen die gruppierten Transaktionsdaten direkt, Cluster- und Entscheidungsbaumanalysen dagegen Transaktionsdaten nur in Form aufbereiteter Datenmodelle verarbeiten können.⁷⁵

3.2.3 Mustererkennung in Webserver Protokolldateien

Im Mittelpunkt der folgenden Beschreibung der Methoden im Data Mining zur Mustererkennung in Webserver Protokolldateien stehen solche Verfahren, die eine automatische Generierung von Hypothesen in Bezug auf die Identifikation individueller Verhaltensmuster erlauben. Dazu gehören die Assoziationsanalyse und die Clusteranalyse; Nicht betrachtet werden dagegen künstliche neuronale Netze (KNN), Sequenzanalysen und Entscheidungsbauminduktionen.

3.2.3.1 Assoziationsanalyse

Durch Assoziationsanalysen können allgemein inhaltliche Abhängigkeiten zwischen Teilmengen eines Datenbestands identifiziert werden.⁷⁶ Die Beschreibung der Assoziationsregeln erfolgt über Signifikanzen auf der Grundlage der Häufigkeit einer identifizierten Regel, dem sog. "Support Faktor", in Verbindung mit der Zuverlässigkeit bzgl. des Auftretens dieser Regel in der Grundgesamtheit der analysierten Daten, dem sog. "Konfidenzfaktor".⁷⁷ Eine Assoziationsanalyse zur Identifikation von Abhängigkeiten zwischen verschiedenen Seitenabfragen auf einem Webserver erfordert zunächst die Bestimmung gruppiert Navigationspfade einzelner Nutzersitzungen. Über den Vergleich der Ähnlichkeit in den Pfadmustern verschiedener Transaktionen können dann signifikante Abhängigkeiten zwischen den Seitenabfragen dieser Transaktionen gebildet werden. Die Notation dieser Abhängigkeiten erfolgt bspw. bei einer Assoziation von 90% zwischen zwei Seitenabfragen A und B, aufgetreten bei 20% aller Transaktionen, in der Form: $A.html \rightarrow B.html^{0,9;0,2}$

Eine Variante der Assoziationsanalyse bildet die sog. Pfadanalyse, die Transaktionen auf gleichförmige lineare Pfadmuster hinsichtlich der Häufigkeit ihres Auftretens untersucht.⁷⁸ Ein signifikantes Pfadmuster bspw. für eine Seite A, von der mit einer Häufigkeit von 40% in allen Transaktionen immer die Seite B aufgerufen wird, wird formal notiert in der Form: $A.html \rightarrow B.html^{0,4}$

⁷⁴ Vgl. Wu/Yu/Ballmann (1998): S. 92 f.

⁷⁵ Vgl. Bensberg (2001), S. 143

⁷⁶ Vgl. Agrawal/Imielinski/Swami (1993)

⁷⁷ Vgl. Bensberg/Weiß (1999), S. 430

⁷⁸ Vgl. Bensberg (2001), S. 149

3.2.3.2 Clusteranalyse

Clusteranalysen dienen allgemein zur Bestimmung möglichst homogener Gruppen in den Teilmengen eines Datenbestands.⁷⁹ Zur Ermittlung der Ähnlichkeit zwischen den Teilmengen ist zunächst die Bestimmung eines Proximitätsmaßes notwendig, über das dann im nächsten Schritt die einzelnen Daten zu Gruppen fusioniert werden. Metrisch skalierte Daten können dabei unmittelbar über die quantitative Distanz ihrer Werte verglichen werden. Ein Beispiel für ein Distanzmaß für metrisch skalierte Daten bildet die Minowski-Metrik in Abhängigkeit eines Distanzgewichtungsfaktors r .⁸⁰ Je höher r gewählt wird, desto höher wird die Differenz zwischen verschiedenen Seitenaufrufen x_i und x_j eines Nutzers k gewichtet und damit der Abstand zwischen Gruppen ähnlicher Distanzwerte größer:

$$d_{ij} = \left[\sum_{k=1}^{n_M} |x_{ik} - x_{jk}|^r \right]^{\frac{1}{r}}$$

Die protokollierten Aktivitäten eines Webservers werden demgegenüber jedoch als nominalskalierte Daten gespeichert. Um ihre Ähnlichkeit über verschiedene Merkmalsausprägungen zu quantifizieren, muss zuvor eine Binärcodierung für diese Merkmale wie dargestellt in Abbildung 13 vorgenommen werden.⁸¹

		Objekt j	
		Eigenschaft vorhanden (1)	Eigenschaft nicht vorh. (0)
Objekt i	Eigenschaft vorhanden (1)	a	c
	Eigenschaft nicht vorh. (0)	b	d

Abbildung 12: Binärcodierung zur Ähnlichkeitsbestimmung⁸²

Anhand der Häufigkeiten der Merkmalsübereinstimmungen bzw. -abweichungen in den Ausprägungen a-d wird die Ähnlichkeit zweier Objekte i und j anhand einer Ähnlichkeitsfunktion s bestimmt. Die Werte λ und δ repräsentieren Gewichtungsfaktoren für die Bedeutung fehlender gemeinsamer Merkmale für die Objektähnlichkeit:

$$s_{ij} = \frac{a + \delta \cdot d}{a + \delta \cdot d + \lambda \cdot (b + c)}$$

3.2.4 Bewertung des Web Usage Mining für die individuelle Nutzeridentifikation

Der Vorteil der Verwendung auf der Basis von Beobachtungen abgeleiteter Verhaltensmuster gegenüber registrierten Präferenz- und Kundenprofilen liegt einerseits in der Lückenlosigkeit der Beobachtbarkeit der Nutzerverhaltens, zum anderen in der Flexibilität in Bezug auf die Identifizierbarkeit momentaner individueller Bedürfnisse. Im Falle des Fehlens einer Benutzer- und Authentifizie-

⁷⁹ Vgl. Bensberg (2001) – evtl. andere Quelle...

⁸⁰ Quelle?

⁸¹ Vgl. Bensberg (2001), S. 115 f.

⁸² In Anlehnung an Bensberg (2001), S. 116

rungskennung ist demgegenüber jedoch allein aus der Verhaltensbeobachtung keine eindeutige Identifizierbarkeit gewährleistet. Die konsolidierten Verhaltensmuster entsprechend der dargestellten Methoden ermöglichen lediglich eine Kategorisierung individueller Aktivitäten in Gruppen typischer Ähnlichkeiten und erlauben entsprechend auch keine eindeutige Reidentifikation eines bestimmten Nutzers im Falle unterschiedlichen Nutzungsverhaltens im Zuge verschiedener Sitzungen.

Die nahe liegende Lösung des Problems der eindeutigen Identifizierbarkeit bei gleichzeitiger Wahrung der situativen Flexibilität, die Kopplung zwischen personenbezogenen und verhaltensbezogenen Daten unterliegt wiederum wesentlichen rechtlichen Restriktionen. Die derzeitigen rechtlichen Rahmenbedingungen für die Verwendung von personenbezogenen Daten durch kommerzielle Anbieter wird deshalb im Folgenden kurz dargestellt.

3.3 Rechtliche Grundlagen der Nutzung personenbezogener Daten

Die Verarbeitung personenbezogener Daten durch sowohl öffentliche Stellen von Bund und Ländern als auch durch nicht-öffentliche Stellen wie privatwirtschaftliche Unternehmen ist in Deutschland reglementiert durch das Bundesdatenschutzgesetz (§ 1, Abs. 2 BDSG). Das BDSG dient dem Schutz des Einzelnen davor, "dass er durch den Umgang mit seinen personenbezogenen Daten in seinem Persönlichkeitsrecht beeinträchtigt wird." (§ 1, Abs. 2 BDSG). Als personenbezogene Daten gelten dabei alle Einzelangaben "über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbarer Person" (§ 3 BDSG).

Als spezielle Rechtsnorm wird das Datenschutzrecht seit 1997 durch das Gesetz über die Nutzung von Telediensten (TDG) und das Teledienste Datenschutzgesetz (TDDSG) für das Internet konkretisiert. Als Teledienste gelten solche elektronischen Informations- und Kommunikationsdienste, "die für eine individuelle Nutzung von kombinierbaren Daten wie Zeichen, Bilder oder Töne bestimmt sind, und denen eine Übermittlung mittels Telekommunikation zugrunde liegt (§ 2, Abs. 1 TDG). Dies betrifft insbesondere Angebote im Bereich der Individualkommunikation wie Telebanking, Datendienste wie bspw. Verkehrs- oder Wetterdaten, Angebote zur Nutzung des Internets, zur Nutzung von Telespielen oder Transaktionsformen für Waren und Dienstleistungen im e-commerce (§ 2, Abs. 2 TDG).

Für die Erhebung, Verarbeitung und Nutzung personenbezogener Daten legen BDSG (§ 4, Abs. 1) und TDDSG (§ 3, Abs.1) ein Verbot mit Erlaubnisvorbehalt fest: Ausgenommen einer ausdrücklichen Erlaubnis durch andere Rechtsnormen dürfen personenbezogene Daten nur im Falle einer Einwilligung des Nutzers erhoben, verarbeitet und vom Dienstleister genutzt werden (§ 3, Abs. 2 TDDSG). Die Einwilligung durch den Nutzer muss als eindeutige, bewusste Handlung erfolgen, als solche vom Dienstleister protokolliert und vom Nutzer sowohl jederzeit abgerufen als auch widerrufen werden können (§ 4, Abs. 2 TDDSG). In der Gesetzesfassung von 2001 hat der Anbieter die Pflicht, den Nutzer zwar über Art, Umfang und Zweck der Erhebung, Verarbeitung und Nutzung seiner Daten zu unterrichten und seine Einwilligung abzuholen, jedoch erst ab Beginn der Nutzung dieser Daten im Rahmen seines Dienstangebots (§ 4, Abs. 1 TDDSG).

Hinsichtlich des Verwendungszwecks personenbezogener Daten unterscheidet das TDDSG explizit zwischen Bestands- und Nutzungsdaten. Bestandsdaten sind personenbezogene Daten, die für die Begründung, inhaltliche Ausgestaltung und Änderung eines Vertragsverhältnisses über die Nutzung von Telediensten erforderlich sind (§ 5 TDDSG). Bestandsdaten dürfen zweckgebunden auch ohne Einwilligung des Nutzers erhoben, verarbeitet und genutzt werden, ihre Verwendung zum Zwecke jeglicher Form der Marktforschung bedarf jedoch der Einwilligung. Nutzungsdaten umfassen dagegen solche Daten, die zur Ermöglichung der Inanspruchnahme und der Abrechnung von Telediensten dienen. Dazu gehören insbesondere Merkmale zur Identifikation eines Nutzers, Angaben über Beginn, Ende und Umfang der Nutzung sowie Angaben über die in Anspruch genommenen Teledienste (§ 6, Abs. 1 TDDSG). Nutzungsdaten dürfen zum Zwecke von Inanspruchnahme und Abrechnung eines Dienstes auch ohne Einwilligung des Nutzers erhoben und bis zu einer Frist von 6 Monaten gespeichert werden. Für solche Daten, die nicht zur Abrechnung dienen, besteht jedoch über das Ende des Nutzungsvorgangs hinaus ein Löschgebot (§ 6, Abs. 4 TDDSG).

Ein Beispiel für abrechnungsbezogene Nutzungsdaten, die ohne Einwilligung des Nutzers erhoben und über den Nutzungsvorgang hinaus gespeichert werden dürfen, bilden die im oben genannten Beispiel des Buchhändlers Amazon.de zur Transaktionsabwicklung notwendigen Zahlungsangaben. Die im voranstehenden Abschnitt beschriebenen Protokolldaten dürfen dagegen nur im Rahmen der Dauer der Inanspruchnahme eines Dienstes erfasst werden. Eine Speicherung über die Nutzungsdauer hinaus bedarf dagegen der Anonymisierung der Daten. Dabei gilt ein explizites Kopplungsverbot zwischen allgemeinen Einzelangaben und solchen Merkmalen, die mit denen Einzelangaben einer bestimmten oder bestimmaren natürlichen Person zugeordnet werden können.

4 Produktindividualisierung

Wie bereits eingangs von Kapitel 2 beschrieben, erwirbt ein Käufer genau dasjenige Produkt, für das er zum Kaufzeitpunkt die höchste Präferenz empfindet.⁸³ Aufgrund der beschriebenen Identifizierbarkeit individueller Merkmale im Rahmen eines rechnergestützten Dialogs auf der Grundlage interaktiver Medien wie des Internets eröffnen sich dabei Möglichkeiten, den Identifikationsvorteil gegenüber dem Kunden in einer individuellen Gestaltung des Produktangebots auszunutzen und dadurch den Kaufanreiz für ihn zu erhöhen.

In der Literatur finden sich derzeit, vorwiegend unter dem Begriff "Personalisierung", zwei verschiedene und zum Teil nicht trennscharf verwendete Herangehensweisen an das Thema der Produktindividualisierung: eine Marketingtheoretische und eine Industriebetriebliche. Die marketingtheoretische Sichtweise nähert sich dem Thema der Individualisierung vor dem Hintergrund des Auswahlproblems eines Kunden in den heute von Überfluss und Überangebot geprägten Märkten. Eine individuelle Ausrichtung der Marketinginstrumente bedeutet daher den Aufbau einer individuellen Kundenkommunikation zur Verringerung dieses Auswahlproblems.⁸⁴ Als Mittel dienen dazu im Rahmen der Produkt-

⁸³ Vgl. Runte (2000), S. 1

⁸⁴ Vgl. Runte (2000), S. 5

und Kommunikationspolitik persönliche, auf die Bedürfnisse des Kunden zugeschnittene Produktempfehlungen. Aus Sicht der Industriebetriebslehre besteht demgegenüber das zentrale Hindernis einer wettbewerbsstrategisch wünschenswerten individuellen Differenzierung des Produktangebots im traditionellen Dilemma durch den Kostenanstieg im Zuge der Erhöhung der Variantenzahl bei gleich bleibender Produktionsmenge.⁸⁵ Moderne IuK-Technologien verringern jedoch diese Hürde einerseits aufgrund einer besseren innerbetrieblichen Koordinationseffizienz, andererseits aufgrund der bereits beschriebenen Verbesserung hinsichtlich der Identifizierbarkeit und datentechnischen Integrierbarkeit von Kundenwünschen.⁸⁶ Eine effizient koordinierte und an den Kundenbedürfnissen flexibel orientierte modulare Produktionsstruktur erlaubt somit den sinnbildlichen Spagat "zwischen den Stühlen" von Differenzierung und Kosteneffizienz in Form einer individuellen Konfiguration flexibler Produktmodule (oder "*Mass Customization*") zu einem Kostenniveau, das dem der reinen Massenfertigung entspricht.⁸⁷

Marketing- und industriebetriebliche Sichtweise sind grundsätzlich komplementär, insofern neben Formen personalisierter Produktempfehlungen auch individuelle Güterkonfigurationen im Sinne der *Mass Customization* auch als Instrumente einer individuellen Produktpolitik im Rahmen des Marketingmix interpretiert werden können. Im Folgenden werden deshalb zunächst allgemein verschiedene Methoden für ein individualisiertes Produktangebot unabhängig vom Personalisierungstyp vorgestellt. Im zweiten Abschnitt wird dann anhand einer Betrachtung praktischer Umsetzungsbeispiele von Individualisierungssystemen eine Differenzierung in Bezug auf die Tiefe der individuellen Konfiguration eines Produkts, zwischen im obigen Sinne individuellen Empfehlungssystemen und individuellen Produktkonfigurationssystemen, vorgenommen sowie ein Ausblick auf in diesem Zusammenhang offene Forschungsfragen gegeben.

⁸⁵ Vgl. Porter (1999), S. 79

⁸⁶ Vgl. Piller (2001), S. 226 f. und Picot/Reichwald/Wigand (1998), S. 91

⁸⁷ Vgl. Piller (2001), S. 200

4.1 Individuelle Empfehlungssysteme

Den Kern eines Personalisierungssystems bildet ein sog. Individualisierungs-Regel-System, das die Art und Weise festlegt, wie Informationen und Angebote dem Kunden präsentiert werden.⁸⁸ Ein solches System besteht im Wesentlichen aus drei Bestandteilen: einem Nutzermodell, in dem die Präferenzen eines Nutzers auf der Grundlage demographischer Angaben oder verhaltensbezogener Beobachtungen abgeleitet werden; einem Ressourcenmodell, das Produktmerkmale enthält die auf der Grundlage von Produkteigenschaften abgeleitet werden; einem Inferenzmechanismus, ein automatischen Verfahren, das die individuelle Zuordnung zwischen Ressourcen und Kunden vornimmt (vgl. Abbildung 14).

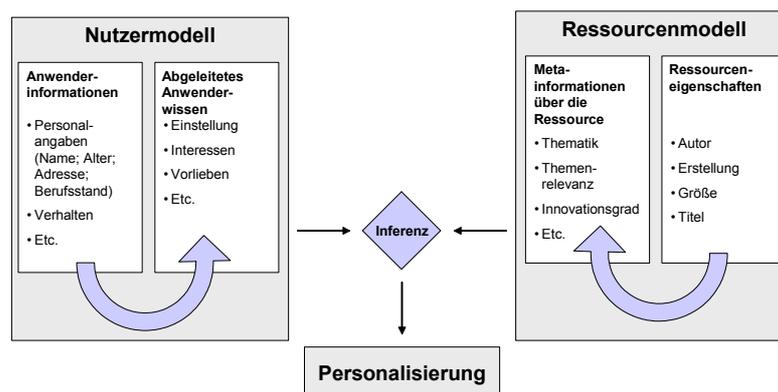


Abbildung 13: Individualisierungs-Regel-System⁸⁹

Grimm et al. unterscheiden allgemein fünf Verfahrensweisen für eine Individualisierung des Produktangebots auf der Grundlage interaktiver Medien (vgl. Abbildung 15):⁹⁰ die Produktauswahl durch explizite Vorgaben des Nutzers, entweder durch Selbstselektion des Nutzers oder auf Basis automatisierter eigenschaftsbasierter Filterregeln für das Produktangebot durch Präferenzprofile; die Produktempfehlung auf Grundlage signifikanter Übereinstimmungen zwischen verschiedenen Kunden durch Collaborative Filtering; das automatische Filtern von Produkteigenschaften durch beobachtungs-basierte Kundenprofilen; die Auswahlentscheidung durch intelligente Agentensysteme, die entweder direkt auf dem Endgerät des Kunden oder durch Aktivierung durch das Kundenendgerät auf dem Server des Anbieters ausgeführt werden.

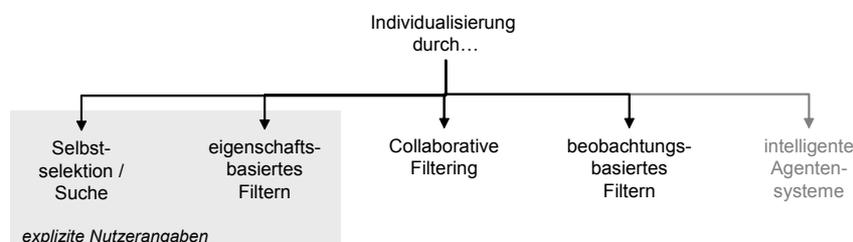


Abbildung 14: Verfahren zur Produktindividualisierung

⁸⁸ Vgl. Grimm/Jüstel/Klotz (2002), S. 181

⁸⁹ In Anlehnung an Schackmann/Link (2001)

⁹⁰ Vgl. Grimm/Jüstel/Klotz (2002), S. 181 ff.

Als intelligente Agenten werden Softwaresysteme bezeichnet, die ihre Aufgabe – entweder auf Initiative des Nutzers hin oder automatisch durch ein bestimmtes Ereignis ausgelöst – weitgehend autonom und zielgerichtet erfüllen. Intelligente Agenten stellen somit streng genommen keine eigene Klasse von Personalisierungssystemen dar, sondern vielmehr eine softwaretechnische Implementierungsform verschiedener Personalisierungsalgorithmen und werden deshalb im Weiteren vernachlässigt.

Der wesentliche Unterschied zwischen den übrigen vier Verfahren liegt in der Konkretisierung der individuellen Präferenzmerkmale im Rahmen der Nutzermodellierung, als Selektionskriterium für die Auswahl des personalisierten Produktangebots. Im Rahmen von Selbstselektion und eigenschaftsbasiertem Filtern werden relevante Produktmerkmale direkt durch in der Regel explizit von einem Nutzer selbst bestimmten Eigenschaftspräferenzen in Bezug auf seinen Produktwunsch ausgewählt. Collaborative Filtering Methoden identifizieren demgegenüber systematische Ähnlichkeiten in den Merkmalen verschiedener Nutzer und bilden darüber Annahmen über analoge Produktpräferenzen. Beobachtungsbasierte Regelsysteme ordnen schließlich Aktivitäten spezifischen Typologien gespeicherter Verhaltensmuster und den diesen Mustern beigeordneten Empfehlungen zu. Im Folgenden wird der derzeitige Forschungsstand in Bezug auf Individualisierungssysteme entsprechend dieser Methodensystematik anhand ausgewählter Theoriebeispiele vorgestellt. Am Ende des Abschnitts erfolgt eine Gegenüberstellung der verschiedenen Methoden und eine Bewertung ihrer Spezifika.

4.1.1 Individualisierung durch explizite Nutzerangaben

Der einfachste Fall einer Individualisierung des Produktangebots bildet die Selbstselektion des Nutzers für ihn relevanter Produktmerkmale oder -eigenschaften. Selektionsgrundlage bilden Auswahllisten, die der Anbieter entweder fakultativ im Rahmen eines personalisierten Zusatzdienstes oder obligatorisch für die Nutzung des gesamten Dienstangebots aus der Gesamtheit der möglichen Produktmerkmale und -eigenschaften bereitstellt. Die jeweilige Auswahl eines Nutzers wird in einem registrierten Profil gespeichert. Auf der Grundlage dieses Präferenzprofils wird die Produktauswahl beim nächsten Dienstauftrag des Nutzers auf diejenigen Angebote mit den entsprechenden Eigenschaften und Merkmalen reduziert. Abbildung 6 in Kapitel 3.1.1 zeigt einen Ausschnitt einer solchen Auswahlliste am Beispiel von Amazon.de.

Eine Variante der Selbstselektion bildet das Verfahren des sog. "eigenschaftsbasierten Filterns", das eine Angebotsindividualisierung lediglich auf der Grundlage der sprachlichen Produkteigenschaften, jedoch ohne Modellierung spezifischer inhaltlicher Merkmale vornimmt. Ziel des eigenschaftsbasierten Filterns bildet die Ableitung einer Klassifikationsregel für vom Nutzer ungesehene Objekte hinsichtlich ihrer Passgenauigkeit zu seinen Präferenzen anhand von signifikanten Korrelationen zwischen Produktinhalt und Präferenzprofil.⁹¹

Van Meteren und Van Someren stellen eine solche Klassifikationsregel am Beispiel eines Internetdienstes für Heimwerkerinformationen vor:⁹² Ein in diesem Fall rein textbasiertes Produkt X kann

⁹¹ Vgl. van Meteren/van Someren (2000) und Runte (2000), S. 10 f.

⁹² Vgl. van Meteren/van Someren (2000)

allgemein beschrieben werden als Vektor \vec{x} seiner sprachlichen Merkmale im n-dimensionalen Merkmalsraum in der Form $\vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ bzw. als Ortsvektor durch den Punkt $X = (x_1, \dots, x_n)$. Zur Reduktion des

Merkmalsraums auf signifikant auffällige Spracheigenschaften wird im Rahmen der Datenvorbereitung die Löschung irrelevanter Textbausteine wie Stoppwörter oder HTML-Tags sowie die Textreduktion auf Wortstämme vorgeschlagen. Als Ergebnis dieser Reduktion kann ein Dokument D beschrieben werden als gewichteter m-dimensionaler Vektor $D = (w_1 t_1, \dots, w_m t_m)$ seiner signifikant häufigen Wortstammmerkmale t_i . Der Gewichtungsfaktor w_i kann dabei über das sog. *tf-idf* Schema⁹³ anhand der Merkmalshäufigkeiten tf_i bestimmt werden durch $w_i = tf_i \cdot \log\left(\frac{n}{df_i}\right)$, wobei n die Anzahl der Dokumente im Produktangebot und df_i die Anzahl von Dokumenten mit mindestens einem Merkmal t_i darstellen.

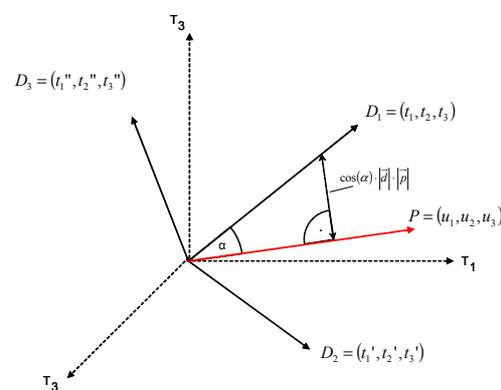


Abbildung 15: Ähnlichkeitsbestimmung durch eigenschaftsbasiertes Filtern⁹⁴

Das Präferenzprofil eines Nutzers kann analog zur Vektorisierung der Produktinhalte als Vektor präferierter Merkmalsausprägungen durch den Punkt $P = (u_1, \dots, u_n)$ abgebildet werden, wobei u_i die jeweils präferierte Ausprägung des Nutzers in Bezug auf das Merkmal i darstellt. Der Vektor \vec{p} wird allgemein bezeichnet als "Rating" eines Nutzers über die betrachteten Merkmalsdimensionen.⁹⁵ Das Präferenzprofil wird anhand von Trainingsdaten mit beispielhaften Dokumenten generiert, zu denen der Nutzer eine Bewertung abgeben muss. Damit kann schließlich die Ähnlichkeit eines Dokuments mit den Präferenzen des Nutzers anhand der Vektorähnlichkeit über den geometrischen Abstand bestimmt werden. Eine solche Ähnlichkeitsmetrik bildet das sog. "Cosinus-Maß" für die Verhältnis zweier Vektoren mit dem Zwischenwinkel α in Bezug auf die betrachteten Merkmalsdimensionen, wie als Beispiel für einen dreidimensionalen Merkmalsraum dargestellt in Abbildung 16.⁹⁶ Das Maß für die Ähnlichkeit $sim(D, P)$ zwischen einem Dokument D und dem Präferenzprofil P eines Nutzers kann berechnet

⁹³ term-frequency – inverse document frequency

⁹⁴ In Anlehnung an Salton/Wong/Yang (1975), S. 614

⁹⁵ Vgl. Runte (2000), S: 35 f.

⁹⁶ Wenn Vektoren ähnlich ausgerichtet sind, ist der Winkel zwischen ihnen entsprechend klein und der Cosinus des Winkels bewegt sich gegen eins. Sind zwei Vektoren genau gleich ist der Winkel zwischen ihnen 0 und der Cosinus genau 1.

werden über das Skalarprodukt der beiden Vektoren, auf den Einheitskreis normiert um das Produkt

ihrer Beträge⁹⁷ in der Form:
$$\text{sim}(D, P) = \cos(\alpha) = \frac{D \cdot P}{\|D\| \cdot \|P\|} = \frac{\sum_k u_k \cdot w_k}{\sqrt{\sum_k u_k^2 \cdot \sum_k w_k^2}}$$

4.1.2 Individualisierung durch Collaborative Filtering

Die Verwendung des eigenschaftsbasierten Filterverfahrens zur Individualisierung des Produktangebots besitzt in der Praxis zwei Schachstellen:⁹⁸ Erstens kann das Verfahren nur bei solchen Produkten angewandt werden, deren Inhalt einfach rechnergestützt analysierbar ist. Es versagt jedoch bei Produkten, die nur einen geringen Textanteil besitzen, wie Software, Bilder, Ideen oder Meinungen. Zweitens können solche Produkte generell nicht empfohlen werden, die keine semantischen Übereinstimmungen mit dem Präferenzprofil des Nutzers aufweisen. Eine zweite Gruppe von Empfehlungsalgorithmen, sog. "Collaborative Filtering" (CF) Verfahren, verwenden daher, anstelle einer direkten Ähnlichkeitsüberprüfung zwischen Produkteigenschaften und Nutzerpräferenzen, strukturelle Ähnlichkeiten zwischen den Präferenzprofilen verschiedener Nutzer als Grundlage einer individuellen Produktempfehlung.

Ebenso wie eigenschaftsbasierte Filterverfahren verwenden CF-Algorithmen gespeicherte Produktbewertungen als Präferenzprofile. Grundlage des Filterverfahrens bildet die Datenmatrix (U) sämtlicher Ratingvektoren ($u_1 \dots u_M$) relevanter Nutzer (M) innerhalb der betrachteten Gruppe von Objekten (N).

$$U = \begin{pmatrix} u_1 \\ \dots \\ u_i \\ \dots \\ u_M \end{pmatrix} = \begin{bmatrix} u_{11} & \dots & u_{1j} & \dots & u_{1N} \\ \dots & \dots & \dots & \dots & \dots \\ u_{i1} & \dots & u_{ij} & \dots & u_{iN} \\ \dots & \dots & \dots & \dots & \dots \\ u_{M1} & \dots & u_{Mj} & \dots & u_{MN} \end{bmatrix}$$

Abbildung 16: Datenmatrix der Ratingpräferenzen beim Collaborative Filtering

In der Realität ist die Datenmatrix aufgrund des bereits angesprochenen Informationsproblems hinsichtlich der Erhebung von Nutzerpräferenzen unvollständig. Ziel der CF-Methodik besteht deshalb darin, die fehlenden Werte im Rating-Vektor jeweils eines Kunden über Ähnlichkeiten zu anderen in der Datenmatrix gespeicherten Ratings durch Prognosewerte zu ersetzen. Die Ableitung von individuellen Produktempfehlungen umfasst beim CF drei Prozessschritte:⁹⁹

⁹⁷ Vgl. Hess/Schneider (2003): Der Abstand zwischen zwei Punkten $D_1 = (t_1, t_2, t_3)$ und $P = (u_1, u_2, u_3)$ ist definiert durch ihren sog. "euklidischen Abstand" als $\sqrt{(u_1 - t_1)^2 + (u_2 - t_2)^2 + (u_3 - t_3)^2}$. Darüber hinaus ist das Skalarprodukt (oder innere Produkt) der Ortsvektoren dieser Punkte proportional zum Cosinus des Winkels zwischen den beiden Produkten bzw. gleich dem Produkt von Cosinus des Zwischenwinkels und den Beträgen der beiden Vektoren: $\vec{d} \cdot \vec{p} = \cos(\alpha) \cdot |\vec{d}| \cdot |\vec{p}|$. Das Skalarprodukt wird häufig zur Abstandsbestimmung vorgezogen, das es sich in der Regel leichter berechnen lässt.

⁹⁸ Vgl. Melville/Mooney/Nagarajan (2001), S. 1

⁹⁹ Vgl. Melville/Mooney/Nagarajan (2001), S. 3 f.

- Systematische Ähnlichkeiten zwischen verschiedenen Kunden werden anhand von Distanz- oder Korrelationsanalysen identifiziert. Eine der gebräuchlichsten Maßzahlen, um für – wie in diesem Fall – metrische Daten die Stärke des Zusammenhanges zu ermitteln, ist der Korrelationskoeffizient nach Bravais-Pearson. Der Koeffizient der Präferenzen zweier Nutzer a und u , mit den durchschnittlichen Standardabweichungen s_a und s_u von den mittleren Produktbewertungen in der Datenmatrix, ist definiert als: $P_{a,u} = \frac{\text{cov}(a,u)}{s_a \cdot s_u}$. Die Ähnlichkeit zwischen den Nutzerprofilen in der Datenmatrix kann daher berechnet werden in der Form:

$$P_{a,u} = \frac{\sum_{i=1}^m (r_{a,i} - \bar{r}_a) \cdot (r_{u,i} - \bar{r}_u)}{\sqrt{\sum_{i=1}^m (r_{a,i} - \bar{r}_a)^2 \cdot \sum_{i=1}^m (r_{u,i} - \bar{r}_u)^2}}$$

- Ähnliche Benutzer N werden als "Mentoren" bzw. "Nachbarn" selektiert und die jeweils fehlenden Werte der Rating-Vektoren ergänzt.
- Die konkrete Angebotsindividualisierung erfolgt schließlich durch eine Auswahl derjenigen Objekte mit den höchsten individuellen Prognosewerten hinsichtlich der Passgenauigkeit einer analogen Produktempfehlung. Die Prognosewerte eines Produkts i werden berechnet als gewichteter Durchschnitt der Abweichungen der Präferenzen eines Nutzers a vom Mittelwert des Nachbarn u :

$$p_{a,i} = \bar{r}_a + \frac{\sum_{u=1}^n (r_{u,i} - \bar{r}_u) \cdot P_{a,u}}{\sum_{u=1}^n P_{a,u}}$$

Die Genauigkeit der Prognosewerte eines CF-Systems ist zum einen in hohem Maße abhängig von der Anzahl der Benutzer, die das System verwenden, zum anderen von der Anzahl verfügbarer Produktbewertungen. Probleme ergeben sich jedoch bspw., wenn das Verhältnis von Produkten und Nutzern sehr groß ist, wodurch die Wahrscheinlichkeit der Existenz signifikant ähnlicher Bewertungsprofile sinkt. Eine weitere Schwachstelle bildet das sog. "Erstbewerterproblem"¹⁰⁰: Produkte können überhaupt dann erst individuell empfohlen werden, wenn sie von mindestens einem Nutzer bewertet wurden.¹⁰¹

Zur Erhöhung der Prognosegenauigkeit von CF-Verfahren vor dem Hintergrund dieser Probleme schlagen Melville/Mooney/Nagarajan (2001) am Bsp. eines Kinofilm-Empfehlungssystems die Kopplung von eigenschaftsbasierten und kollaborativen Filtern in Form eines "content-boosted collaborative filtering" (CBCF) vor. Wie in Abbildung 18 dargestellt, wird die lückenhafte Datenmatrix der verfügbaren Produktbewertungen aus den Nutzerprofilen zunächst durch ein eigenschaftsbasiertes Filterverfahren vervollständigt. Als Grundlage für die eigenschaftsbasierte Vervollständigung dient im Beispielfall der kommerzielle Internetdienst InternetMovieDatabase (IMDb), der für jeden Film textba-

¹⁰⁰ "first rater problem"

¹⁰¹ Vgl. Melville/Mooney/Nagarajan (2001), S. 1

sierte Angaben bspw. zu Handlung, Ensemble oder Kritiken bereitstellt. Über die vervollständigte Datenmatrix V , die für jeden Nutzer ein eigenschaftsbasiert gefilterten "Pseudo-Ratingvektor"¹⁰² enthält, wird dann wiederum über CF-Algorithmen eine Empfehlung berechnet.

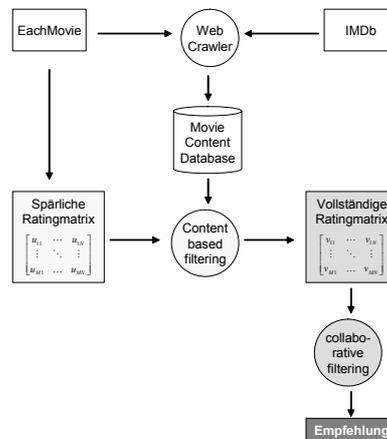


Abbildung 17: Content-boosted Collaborative Filtering

In einer empirischen Testreihe deckten Melville/Mooney/Nagarajan (2001) drei weitere Abhängigkeiten in Bezug auf die Genauigkeit der Empfehlungsergebnisse:

1. Eine Beeinflussung der Korrelation $P_{a,u}$ zwischen den Ratings zweier Nutzer a und u durch die Anzahl n_i der jeweils von beiden überlappend bewerteten Produkte, wobei hoch korrelierte Mentoren mit jedoch nur wenigen überlappenden Produkten das Empfehlungsergebnis negativ beeinträchtigen. Zur Abwertung dieser Korrelationen wird ein Signifikanzgewicht¹⁰³ vorgeschlagen, als:
$$sg_{u,i} = \begin{cases} r_{u,i}, & \text{für } n_i < 50 \\ 1, & \text{für } n_i \geq 50 \end{cases}$$
2. Eine Beeinflussung der Qualität des eigenschaftsbasierten Filterverfahrens durch die Anzahl insgesamt bewerteter Produkte durch einen Nutzer, wobei eine geringe Zahl von Bewertungen durch einen Nutzer a zu einer verfälscht hohen Korrelation zum Profil eines anderen Nutzer u führt. Zur Verringerung der Fehlerwahrscheinlichkeit wird der sog. "harmonisierende Mittelwertgewichtungsfaktor"¹⁰⁴ eingeführt, als:
$$hm_{a,u} = \frac{2m_a m_u}{m_a + m_u},$$
 wobei

$$m_{a/u} = \begin{cases} \frac{n_{a/u}}{50}, & \text{für } n_{a/u} < 50 \\ 1, & \text{für } n_{a/u} \geq 50 \end{cases}.$$

Signifikanzgewicht und harmonisches Mittelwertgewicht beziehen sich beide auf die Korrelation $P_{a,u}$ zwischen den Ratings zweier Nutzer a und u und können daher zusammengefasst werden zum "hybriden Korrelationsgewicht"¹⁰⁵ $hw_{a,u} = hm_{a,u} + sg_{a,u}$.

¹⁰² Die Vektorkoordinaten von V werden notiert als $v_{u,i} = \begin{cases} r_{u,i}, & \text{falls } i \text{ von } u \text{ bewertet} \\ c_{u,i}, & \text{andernfalls} \end{cases}$

¹⁰³ "Significance weighting factor" (sg)

¹⁰⁴ "Harmonic Mean weighting factor" (hm)

¹⁰⁵ "hybrid correlation weight" (hw)

3. Es wird unterstellt, dass die Ergebnisse des eigenschaftsbasierten Filterverfahrens zu Ergänzung der Profillücken, da sie auf den ursprüngliche vom Nutzer spezifizierten Bewertungen beruhen, insgesamt eine höhere Übereinstimmung mit seinen Präferenzen als kollaborative, im Vergleich mit anderen Nutzern ermittelte Prognosewerte. Daher werden die über eigenschaftsbasiertes Filtern ermittelten Koordinaten im Pseudo-Rating eines Nutzers gegenüber Mentorenpräferenzen im Rahmen der Prognosewertberechnung höher gewichtet, durch den sog. "Selbstgewichtungsfaktor": $sg_{u,i} = \begin{cases} r_{u,i}, & \text{für } n_i < 50 \\ 1, & \text{für } n_i \geq 50 \end{cases}$

Unter Einbeziehung der drei Gewichtungsfaktoren lässt sich somit eine Empfehlung für den Nutzer a für ein Produkt i berechnen als:

$$p_{a,i} = \bar{v}_a + \frac{sw_a (c_{a,i} - \bar{v}_a) + \sum_{u=1}^n (v_{u,i} - \bar{v}_u) \cdot hw_{a,u} P_{a,u}}{sw_a + \sum_{u=1}^n hw_{a,u} P_{a,u}}$$

4.1.3 Individualisierung durch beobachtungsbasiertes Filtern

Der Vorteil der Verwendung expliziter Präferenzprofile in Form von nutzerspezifischen Produktbewertungen im Rahmen der Produktindividualisierung besteht darin, dass die Profildaten bereits als metrisch skalierte Produktgewichtungen vorliegen. Jedoch ist wie gezeigt die Qualität der Prognosewerte eines Empfehlungssystems im hohen Maße vom Umfang der angegebenen Produktpräferenzen im Verhältnis zum Umfang des Produktkatalogs abhängig. Wie jedoch bereits erörtert, ist die Bereitschaft von Nutzern aktiv Informationen über sich preiszugeben in der Regel sehr gering. Wie Runte feststellt, setzen sowohl eigenschaftsbasierte als auch kollaborative Filterverfahren nicht, wie in den gezeigten Beispielen, zwangsläufig die Existenz explizit spezifizierter Präferenzen voraus, sondern können auch in Verbindung mit impliziten, aus Verhaltensbeobachtungen abgeleiteten Nutzerprofilen angewendet werden.¹⁰⁶

Ein beobachtungsbasiertes Empfehlungssystem besteht, wie in Abbildung 19 dargestellt, in der Regel aus zwei Komponenten: einem Offlinesystem, das über analytische Auswertungen der protokollierten Aktivitäten eines Webservers typische Verhaltensmuster ableitet und speichert; einem Onlinesystem, das das aktive Verhalten eines Nutzers mit den gespeicherten Verhaltensmustern vergleicht und zuordnet.

¹⁰⁶ Vgl. Runte (2000), S. 24

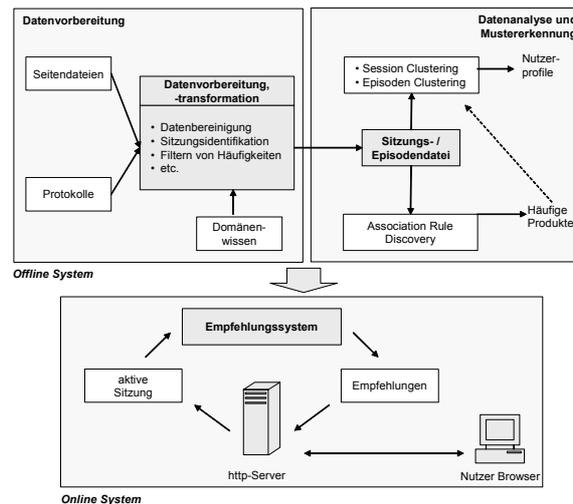


Abbildung 18: Konzept eines beobachtungs-basierten Empfehlungssystems¹⁰⁷

Im Folgenden werden die Verfahren der Profilaggregation im Rahmen der Offlinekomponente und die Ableitung einer nutzerindividuellen Empfehlung im Rahmen der Onlinekomponente am Beispiel eines beobachtungs-basierten Empfehlungssystems vorgestellt.¹⁰⁸

4.1.3.1 Profilaggregation im Rahmen eines beobachtungs-basierten Empfehlungssystems

Wie bereits angesprochen, ist die Auswahl der Verfahren im Rahmen der Datenvorbereitung ebenso wie die Anwendbarkeit des Prognosealgorithmus selbst in hohem Maße abhängig vom gewählten Verfahren zur Mustererkennung von Nutzerprofilen im Rahmen der Datenanalyse. Mobasher/Dai/Luo/Nakagawa (2002) schlagen ein beobachtungs-basiertes Empfehlungssystem vor, das Nutzerprofile in Form von gewichteten Gruppen häufig besuchter URL-Kombination als Grundlage eines eigenschaftsbasierten Filterverfahrens verwendet. Für die Ableitung individueller Nutzerprofile werden zwei Verfahrensweisen vorgeschlagen, die Transaktionsclusterung und die Pageview-Clusterung:

Wie in Abbildung 20 dargestellt, werden im Rahmen der Transaktionsclusterung zunächst – im Zuge der Datenvorbereitung – die von nicht relevanten Aktivitäten gesäuberten Protokolldaten (P) des Webservers durch Pfad- und Assoziationsanalysen zu Cluster individueller Transaktionen und Nutzersitzungen (T) gruppiert (①, ②). Die einzelnen Transaktionen werden als Vektoren gewichteter Objektaufrufe (Pageviews) dargestellt. Als Gewichtungsfaktor können bspw. die Häufigkeit des Aufruf eines Objekts, dessen Betrachtungsdauer oder domänenspezifische Kriterien wie die unterschiedliche Gewichtung zwischen Navigations- und Inhaltsseiten verwendet werden. Die Transaktionen t_i werden anhand der Ähnlichkeit ihrer Pageviews wiederum in Cluster (TC) gruppiert (③). Im letzten Schritt werden diese Transaktionscluster hinsichtlich ihrer Besonderheiten zu individuellen Profilen (pr) aggregiert (④). Dabei wird durch Mittelwertbildung der Gewichtungsfaktoren der Pageviews innerhalb eines Transaktionsclusters diejenigen Seitenaufrufe herausgefiltert, deren Gewichtungsfaktor unter dem jeweiligen Transaktionsmittelwert liegt.

¹⁰⁷ In Anlehnung an Mobasher (2002); Mobasher/Cooley/Srivasta (2000); Mobasher/Dai/Luo/Nakagawa (2001)

¹⁰⁸ Vgl. Mobasher (2002); Mobasher/Cooley/Srivasta (2000); Mobasher/Dai/Luo/Nakagawa (2001)

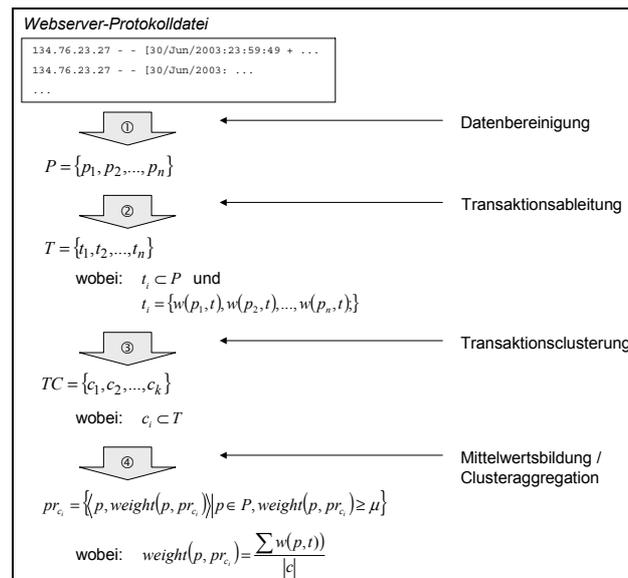


Abbildung 19: Profilaggregation durch Transaktionsclustering

Im Rahmen der Transaktionsclustering werden gemeinsam aufgetretene Pageviews in ähnlichen Transaktionen zu Profilen gruppiert. Im Gegensatz dazu werden im Rahmen der Pageviewclustering häufige Pageviews unabhängig von der Ähnlichkeit der zu Grunde liegenden Transaktionen gruppiert, wodurch auch überlappende Interessen unterschiedlicher Nutzer abgebildet werden können. Dazu wird ein Assoziationsregelverfahren angewendet, dem der Hypergraf der Objekte einer Internetseite, als Repräsentationsform des Beziehungsnetzwerks der Objekte und der sie verbindenden Links, zu Grunde gelegt wird.¹⁰⁹

Wie in Abbildung 21 dargestellt, werden innerhalb der aus den bereinigten Daten (①) aggregierten Transaktionsgruppen T (②) häufig aufgerufene Objekte IS gefiltert (③). Als Maß σ der Häufigkeit eines Seitenaufrufs I_i innerhalb eines Transaktionsclusters c wird die Anzahl der Seitenaufrufe eines Objekts innerhalb des Transaktionsclusters im Verhältnis zur Mächtigkeit sämtlicher Transaktionscluster bestimmt. Diese häufig aufgerufenen Objekte werden im ARHP-Verfahren als Eckpunkte E eines Hypergraphen $H = \langle V, E \rangle$ interpretiert, wobei die Knoten V des Grafen H Elemente der gesäuberten Datenbasis P darstellen. Dabei werden diejenigen Knoten des Hypergraphen herausgefiltert, die eine geringe Verbindung zur den jeweiligen Eckpunkte besitzen. Dazu wird zunächst das Interesse $Interest(I)$ eines individuellen Nutzers an einem häufig aufgerufenen Objekt anhand der Häufigkeit dieses Objektaufrufs im Verhältnis zur Gesamtzahl des Auftretens aller häufigen Objekte I_i in sämtlichen Clustern bestimmt(④).

¹⁰⁹ Association Rule Hypergraf Partinoning (ARHP)

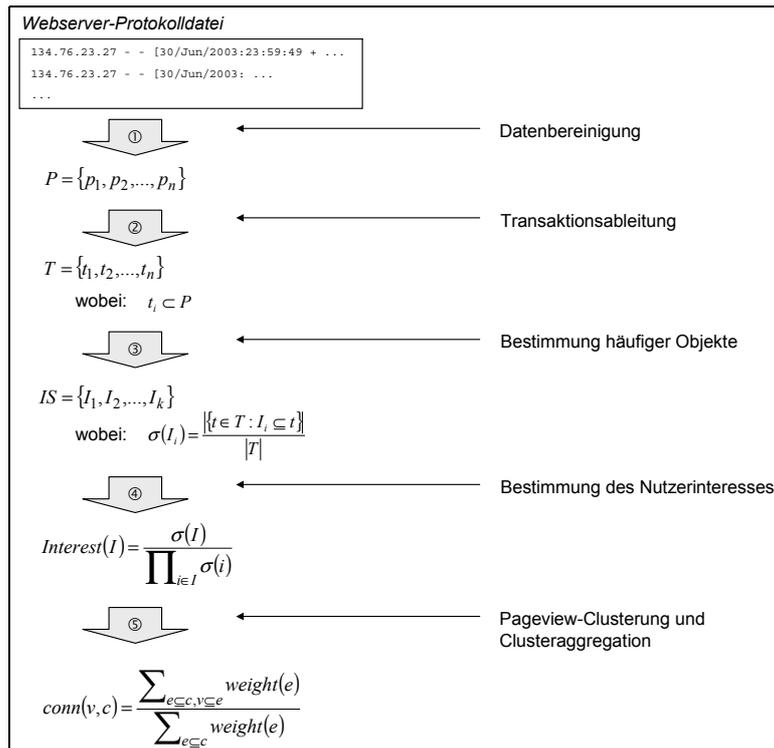


Abbildung 20: Profilaggregation durch Pageviewclustering

4.1.3.2 Ermittlung einer individuellen Empfehlung im Rahmen eines beobachtungs-basierten Empfehlungssystems

Die Onlinekomponente eines Empfehlungssystems zeichnet für die aktive Sitzung eines Nutzers die zuletzt besuchten Seiten über ein Sitzungskontrollprogramm auf. Beispielsweise können die aufgezeichneten Seiteneinträge in einem n -zeiligen verborgenen Fensterbereich protokolliert werden. Anhand von Seiteneigenschaften, wie der Unterscheidung zwischen Inhalts- und Navigationsseiten, kann jeder der besuchten Seiten ein Signifikanzgewicht in Bezug auf die vom Seitenbetreiber für eine Nutzersitzung als bedeutsam erachteten Seiteneigenschaften zugeordnet werden. Ebenso wie die gespeicherten Profile der Offlinekomponente kann damit die aktive Nutzersitzung als n -dimensionaler Vektor $S = \{s_1, s_2, \dots, s_n\}$ der Gewichtungsfaktoren s_i eines aufgerufenen Objekts p_i abgebildet werden,

wobei $S_i = \begin{cases} 1 + s_i & \text{für } p_i \in S \\ 0 & \text{sonst} \end{cases}$. Die entsprechenden n -dimensionalen verfügbaren Profile sind wie voran

stehend beschrieben bestimmt als $C = \{w_1^c, w_2^c, \dots, w_n^c\}$, wobei $w_i^c = \begin{cases} \text{weight}(p_i, C) & \text{für } p_i \in C \\ 0 & \text{sonst} \end{cases}$. Zur Auswahl

des der aktiven Sitzung ähnlichste Profil wird das bereits bekannte Cosinus-Maß verwendet:

$\text{sim}(S, C) = \frac{\sum_k w_k^c \cdot s_k}{\sqrt{\sum_k s_k^2 \cdot \sum_k w_k^{c2}}}$. Die Auswahl von Seiten aus dem ähnlichsten Nutzungsmuster der Online-

komponente kann abschließend anhand der Gewichtungsfaktoren der Objekte im Profilmuster und dem Übereinstimmungsgrad zwischen Profilmuster und aktiver Sitzung anhand eines Empfehlungs-

grads berechnet werden: $\text{Re } c(S, n) = \sqrt{\text{weight}(n, C) \cdot \text{match}(S, C)}$.

4.1.4 Bewertung der dargestellten Verfahren

Das in Kapitel 4.1.1 erläuterte Verfahren zum eigenschaftsbasierten Filtern von Inhalten entsprechend individueller Präferenzen lässt sich wie dargestellt nur auf solche Inhaltsressourcen anwenden, deren Merkmale automatisch computergestützt analysierbar sind. Es stellt einerseits zwar ein vergleichsweise einfach anzuwendendes Verfahren zur Überprüfung semantischer Ähnlichkeiten zwischen den Präferenzen eines Nutzers und den Merkmalen einer Ressource dar, das (ausgenommen der Wortstammanalyse der Ressourcen) ohne aufwendigen analytischen Vorlauf eingesetzt werden kann. Es setzt andererseits allerdings die Bereitschaft des Nutzers voraus, seine Präferenzen über Produktbewertungen zu spezifizieren. Zudem werden nur solche Produkte von der Filterregel erfasst, deren auffällige Merkmale spezifizierten Präferenzen des Nutzers entsprechen. Solche Ressourcen, die zwar eine sinnvolle sachliche Anknüpfung an die Präferenzen eines Nutzers bieten würden, jedoch keine sprachlichen Gemeinsamkeiten aufweisen, werden von der Filterregel nicht erfasst. Ein Beispiel für einen solchen Fall im Rahmen eines individuell gefilterten Nachrichtenportals wäre ein Bericht über einen Gesetzesentwurf zur Verschärfung des Dopingstrafrechts, der unter Umständen einem sportinteressierten Leser mangels signifikanter sprachlicher Übereinstimmungen mit spezifizierten Präferenzen im Bereich Fußball und Tennis nicht angeboten wird.

Genau hier liegen die Stärken von Collaborative Filtering Verfahren: Anstelle eines direkten Vergleichs zwischen den Eigenschaften eines Produkts und den Präferenzen eines Nutzers werden solche Ressourcen individuell gefiltert, denen andere Nutzer mit grundsätzlich ähnlich strukturierten Präferenzen eine hohe Relevanz beigemessen haben. Dadurch werden unter Umständen auch solche Ressourcen von der Filterregel erfasst, die wie im oben genannten Fall keine signifikanten sprachlichen Gemeinsamkeiten mit den Präferenzen eines Nutzers besitzen aber eventuell einen sachlichen Zusammenhang aufweisen. Dabei können auch solche Ressourcen gefiltert werden, deren Inhalte computergestützt nicht automatisch analysiert werden können. Der Nachteil von Collaborative Filtering Verfahren liegt im langen Vorlauf zum Aufbau eines repräsentativen Umfangs an Referenzprofilen für den Ähnlichkeitsvergleich der Präferenzen eines Nutzers. Problematisch ist der Einsatz von Collaborative Filtering Verfahren zudem bei solchen Angeboten, deren Ressourcen eine hohe Innovationsdynamik aufweisen. Liegen noch keine Bewertungen einer Ressource durch mindestens einen Nutzer vor, ist diese Ressource auch für keinen anderen Nutzer filterbar.

Gegenwärtig werden in der Literatur vorwiegend solche Individualisierungssysteme diskutiert, deren Nutzermodell auf explizit durch den Nutzer spezifizierten Präferenzmerkmalen basieren. Dennoch setzen grundsätzlich wie bereits angesprochen weder eigenschaftsbasierte Filterverfahren noch das Collaborative Filtering ein solches explizites Präferenzprofil voraus. In neueren Veröffentlichungen wird stattdessen zunehmend die Möglichkeit einer Modellierung signifikanter Nutzermerkmale auf der Grundlage beobachtbarer und vom Webserver protokollierter Verhaltensaktivitäten. Neben dem Wegfall des bereits an anderer Stelle mehrfach thematisierten Problems der in der Regel mangelnden Bereitschaft von Nutzern, im Internet persönliche Angaben zu machen, haben beobachtungsgestützten Verfahren den Vorteil, dass nur die aktuellen situative Bedürfnisse eines Nutzers unabhängig von seinen demografischen Angaben oder seinen Präferenzen in der Vergangenheit die Grundlage des Filterverfahrens bilden. Bewegt sich beispielsweise der Nutzer eines Onlinebuchladens mit einer

generellen Vorliebe für klassische Literatur im Rahmen seiner Urlaubsvorbereitung auffällig lange auf den Seiten für Reiseliteratur über eine bestimmte Urlaubsregion, werden unabhängig von seinen sonstigen Vorlieben auch Literaturempfehlungen für die entsprechende Reiseregion von der Filterregel erfasst. Im Vergleich zu den bereits dargestellten Filterverfahren auf Grundlage expliziter Präferenzangaben erhöht sich jedoch durch eine Verwendung beobachtungsbasierter aggregierter Profile die Komplexität des Filterverfahrens und dadurch die Vorlaufdauer für dessen Einsatz. Diese Komplexität wird insbesondere beeinflusst durch die Länge der aufgezeichneten Aktivitäten der jeweils aktuellen Nutzersitzung durch die Onlinekomponente und dadurch die Länge der notwendigen Referenzcluster der Offlinekomponente des Empfehlungssystems. Neben der verfahrensimmanenten Komplexität der Profilaggregation liegt ein weiteres Problem solcher System, die ausschließlich auf der Verwendung beobachtungsbasierter Nutzungsprofile beruhen, in der mangelnden eindeutigen Identifizierbarkeit des Nutzers. Sollen andere Merkmale als die durch das situative Verhalten innerhalb der aktuellen Sitzung rückschließbaren (wie bspw. Alter, Geschlecht, Wohnort, etc.) zur Individualisierung herangezogen werden, muss das beobachtungsbasierte Verfahren zwangsläufig mit expliziten Angaben des Nutzers kombiniert werden.

In Abbildung 21 sind die Merkmale der dargestellten Individualisierungsverfahren nochmals im Überblick zusammengefasst.

Systemtyp	Systemkomponente	Merkmalsausprägung
1. Eigenschaftsbasiertes Filtern	Ressourcenmodellierung	Reduktion einer textbasierten Ressource auf den Vektor ihrer signifikant häufigen Wortstamm-Merkmale
	Nutzermodellierung	Präferenzprofile als vektorisierte Produktratings, erstellt auf der Basis von Trainingsdaten
	Matching	Verhältnisähnlichkeit von Rating- und Ressourcenvektor (Cosinus-Maß)
	Fazit	<ul style="list-style-type: none"> • Rechnergestützte Analysierbarkeit von Inhalten vorausgesetzt • Matching nur bei semantischen Übereinstimmung zwischen Ressourcen- und Nutzerprofilen • Ohne Vorlauf einsetzbar • Gute Empfehlungsqualität für bewertete Produktmerkmale
2. Collaborative Filtering	Ressourcenmodellierung	entfällt
	Nutzermodellierung	Präferenzprofile als vektorisierte Produktratings, erstellt auf der Basis von Trainingsdaten
	Matching	Mentorenselktion anhand von Merkmalskorrelationen und Produktprognose anhand von Mittelwertsvergleichen zwischen Nutzer und Mentor
	Fazit	<ul style="list-style-type: none"> • Startfristproblem und typische Dürftigkeit der Ratingmatrix • First-Rater-Problem • Auch für nicht-rechnergestützt analysierbare Ressourcen • Möglichkeit der Empfehlungen relevanter Ressourcen ohne semantische Profilübereinstimmung
2. Beobachtungsbasiertes Filtern	Ressourcenmodellierung	Vektorisierung der Pageviews der aktuellen Nutzersitzung
	Nutzermodellierung	<ul style="list-style-type: none"> a) Gewichtete Pageview-Transaktionscluster b) Assoziierte überlappende Pageview-Cluster
	Matching	Verhältnisähnlichkeit der Vektoren der aktuellen Sitzung und der abgeleiteten Profile und Auswahl des Objekts mit der höchsten Ähnlichkeitsrelevanz
	Fazit	<ul style="list-style-type: none"> • Zeitlicher analytischer Vorlauf zur Clusterbildung • Abhängigkeit der Empfehlungsqualität von der Aufzeichnungstiefe der aktuellen Sitzung impliziert hohen Rechenaufwand für gute Empfehlungen • Trennung zwischen demografischem Profil und situativen Bedürfnissen • Keine expliziten Profilingaben notwendig

Abbildung 21: Bewertung verschiedener Personalisierungsmethoden

4.2 Praktische Anwendungen von Empfehlungssystemen im Rahmen der Produktindividualisierung

Noch vor zwei Jahren beschieden Zerdick et al. den Möglichkeiten der Personalisierung mehr Versprechen denn Realität der Internetökonomie zu sein.¹¹⁰ Diese Einschätzung trifft heute nicht mehr zu. Eine Reihe von Internetseiten, gleichwohl ob als Distributionskanal für Informationsgüter oder als Bestelloberfläche für physikalische Produkte, bieten die Möglichkeit einer nutzergerechten Individualisierung des Dienstangebots. Im Folgenden werden kurz drei Beispiele typischer Individualisierungsformen im Internet vorgestellt, der Portaldienst "Mein Yahoo!", der Produktkonfigurator "Nike-ID" und Produktempfehlungsdienste im Angebot des virtuellen Buchladens von "Amazon.de". Anschließend werden die in der Praxis verwendeten Personalisierungsformen bewertet und derzeit bestehende Lücken in Bezug auf praktische Implementierungen und theoretische Konzepte aufgezeigt.

4.2.1 Praxisbeispiele für Individualisierungsmöglichkeiten im Internet

Als eines der ersten Unternehmen überhaupt bot Yahoo! eine Personalisierung der Darstellung seiner Portaloberfläche im Internet an. Dabei lassen sich getrennt voneinander die dargestellten Inhaltsklassen auswählen, die Farbeigenschaften der Inhaltspräsentation und die Struktur des Seitenaufbaus festlegen. Die Konfiguration erfolgt, wie in Abbildung 22 dargestellt, anhand von Auswahllisten. Die individuelle, jeweils explizit vom Nutzer bestimmte Auswahl wird einem registrierten Profil zugeordnet und gespeichert. Sein Einverständnis vorausgesetzt, kann ein Nutzer und seine vorgenommenen Konfigurationseinstellungen nach einmaliger Registrierung automatisch über Cookies re-identifiziert werden.

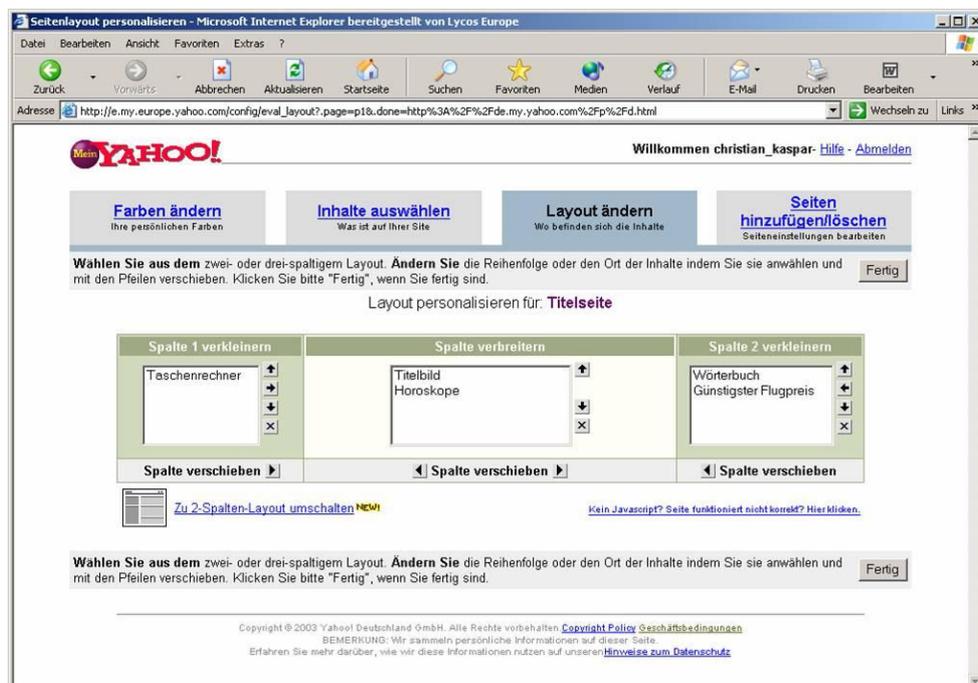


Abbildung 22: Praxisbeispiel 1 - Mein Yahoo!

¹¹⁰ Vgl. Zerdick et al., S. 113

Eine mittlerweile typische Seite für die individuelle Konfiguration physischer Produkte im Sinne der "Mass Customization" bildet der Dienst "Nike-ID" des amerikanischen Sportartikelherstellers Nike, durch den ein Nutzer einen individuellen Turnschuh gestalten kann. Eine Auswahl von 13 Grundtypen der Nike Schuhmodellreihe kann in Bezug auf die Farbgebung der einzelnen Aufbauelemente, wie Sohle, Zunge oder Obermaterial des Überschuhs konfiguriert, mit einem persönlichen Logo individuell signiert und natürlich in der richtigen Größe bestellt werden. Einige Modelle bieten zudem die Möglichkeit, die Weite und die Dämpfungstärke der Sohle zu bestimmen. Abbildung 23 zeigt die Konfigurationsoberfläche in einem Teilschritt des Bestellprozesses. Der Konfigurationsvorgang erzeugt ein digitales Abbild des zu fertigenden individuellen Schuhs, das für die Dauer der Produktion bis zur Versandausführung von Nike gespeichert wird. Der Konfigurationsdienst erfordert keine individuelle Registrierung und wird mit einer verbindlichen Bestellung abgeschlossen. Die Fertigung des individuellen Schuhs selbst erfolgt bei Nike durch eine relativ einfache individuelle Endkombination modular vorgefertigter Komponenten.



Abbildung 23: Praxisbeispiel 2 - Nike-ID

Wie bereits dargestellt, wird bei Amazon.de ein individuelles Präferenzprofil eines Nutzers auf der Grundlage von Bewertungen für Produktkategorien und Einzelprodukte registriert. Ziel der Individualisierung bei Amazon.de bildet demnach weder die Personalisierung der Darstellungsform der präsentierten Produktinhalte, noch die individuelle Konfiguration von Einzelprodukten, sondern eine individuelle Produktauswahl aus dem umfangreichen Katalogbestand an Medienprodukten. Amazon.de bietet eine Reihe verschiedener Empfehlungsdienste an, sowohl einerseits auf der Grundlage des registrierten Präferenzprofils, als auch auf der Grundlage des kurzfristigen Nutzerverhaltens innerhalb des Produktkatalogs. Formen der anhand des Profils personalisierten Empfehlungen bilden bspw. der Dienst "persönliche Empfehlungen", in dem für jede Produktkategorie ein den spezifizierten Präferenzen als am ähnlichsten bestimmtes Produkt angeboten wird, und der Dienst "Neuigkeiten für Sie", der jeweils aktuelle Neuerscheinungen in den als Interessensbereich spezifizierten Produktkategorien. Die zweite Kategorie von Empfehlungsformen richtet sich unmittelbar nach dem aktuell ausgewählten Einzelprodukt. Es werden einerseits solche Produkte empfohlen, die zusätzlich von

anderen Nutzern, nach Erwerb des aktuell angezeigten Produkts gekauft wurden. Zum anderen werden weitere Produkte desselben Autors oder Künstlers aufgelistet.

4.2.2 Bewertung aktueller Individualisierungsformen im Internet

Wie zu Beginn von Kapitel 4 angesprochen, kann das Thema Produktindividualisierung in zweifacher Hinsicht interpretiert werden: als Personalisierung der Vermarktung über digitale Netze angebotener, jedoch nach dem Prinzip der Massenfertigung produzierter Güter oder als individuelle Konfiguration modular erzeugter Güterkomponenten im Sinne einer kundenindividuellen Massenfertigung. Vor diesem Hintergrund können die derzeit verfügbaren Individualisierungsbeispiele zum einen differenziert werden zwischen solchen Diensten, die eine individuelle Produktempfehlung unkonfigurierter Massengüter erlauben und solchen, die eine individuelle Produktkonfiguration ermöglichen. Zum anderen ergibt sich ein weiteres Differenzierungsmerkmal durch die verwendete Filtermethode, entsprechend der aufgezeigten Alternativen. Dadurch ergibt sich eine Klassifikationsmatrix wie dargestellt in Abbildung 24.

	Selbst-selektion / Suche	eigenschaftsbasiertes Filtern	Collaborative Filtering	beobachtungsbasiertes Filtern
<i>Individuelle Produktempfehlungen</i>	Bspw. "mein Yahoo!"	Bspw. Amazon.de "persönliche Empfehlungen"	Bspw. Amazon.de "Kundenrezensionen"	Amazon.de (m. E.) "verwandte Produkte"
<i>Individuelle Komponentenbündel</i>	Bspw. NIKE iD, Porsche Car Configurator	/	/	/

Abbildung 24: Individualisierungsformen und Anwendungsbeispiele

Es wird zum einen deutlich, dass speziell im Beispiel von Amazon.de im Rahmen der verschiedenen Formen implementierter Produktempfehlungen sowohl eigenschaftsbasierte als auch kollaborative Filterverfahren zum Einsatz kommen. Im Wesentlichen wird als Personalisierungsgrundlage das registrierte Präferenzprofil eines Nutzers verwendet. Die Empfehlungsform "verwandte Produkte" stellt eine reduzierte Variante verhaltensbasierten Filterns dar, insofern zwar das aktive Produkt, jedoch keine über diese singuläre Produktbetrachtung hinausgehende Verhaltensdokumentation als Grundlage der Empfehlung verwendet wird.

Demgegenüber beschränken sich zum anderen derzeit individuelle Güterkonfigurationen auf manuelle Konfigurationen für ein digitales Produktabbild durch einen Nutzer im Rahmen des Bestellvorgangs für physische Produkte. Nach Schackmann und Link lassen sich die grundsätzlich für die industrielle Fertigung physischer Produkte formulierten Prinzipien auch für die Erzeugung von Informationsgütern anwenden:¹¹¹ Durch den Einsatz bspw. datenbankgestützter Content Management Systeme und verwendungsneutraler Auszeichnungssprachen wie dem Protokollstandard XML können modular erzeugte Informationsgüter beliebig kombiniert, erweitert und vertieft werden. Die Abgrenzung zwischen individuellen Konfigurationen und individuellen Empfehlungen ist dabei im Rahmen digitaler

¹¹¹ Vgl. Schackmann/Link (2001)

Produkte nicht immer trennscharf, in sofern auch Sammlungen von Empfehlungen in Form von Link-sammlungen eingebettet in eine Internetseite als individuell konfiguriertes Güterbündel interpretiert werden. Der Unterschied wird jedoch deutlich, betrachtet man das Beispiel inhaltsorientierte Medienprodukte wie Nachrichtenportale:

Grundsätzlich stellen Medienprodukte in der Regel Produktbündel von Informationsgütern dar, die traditionell aus Gründen von Größen- und Verbundeffekten des Vertriebs gekoppelt werden. Eine Zeitung repräsentiert bspw. ein Bündel aus redaktionellen Artikeln, Werbe- und Kleinanzeigen. Im Gegensatz zum "linearen" bzw. statischen Aufbau klassischer Medienprodukte können digital erzeugte Inhalte als modulare Konfigurationen von einzelnen Informationsgüterobjekten interpretiert werden. Diese Bündelinterpretation kann nach Rawolle sowohl aus einer Makro- als auch aus einer Mikrostruktur bewertet werden.¹¹² Während die Perspektive der Makrostruktur ein Informationsgut im Sinne geschlossener Medienobjekte, also in Form eines Textes, eines Bildes oder einer Audio-Sequenz als kleinste betrachtete Einheit versteht, repräsentiert die Mikrostruktur die Zerlegung solcher Objekte in einzelne Inhaltselemente. Die Perspektive der Mikrostruktur kann verdeutlicht werden am Beispiel eines einzelnen Zeitungsartikels, interpretierbar als Bündel einer Überschrift, einer Quellenangabe, einem Textkörper und eventuell einem Bild.

Entsprechend erlauben digitale Medienprodukte eine dreifache strukturelle Individualisierungsmöglichkeit: erstens, die interessensspezifische Auswahl verschiedener Produktbündel im Warenkorb, analog zu den von Amazon.de verwendeten Personalisierungsformen; zweitens, die individuelle Konfiguration der jeweiligen Güterbündel auf der inhaltlichen Makroperspektive, im Sinne einer individuellen Empfehlung von einzelnen Informationsgütern; Ein Beispiel für eine individuelle Konfiguration eines Güterbündels auf Makroebene wäre die individuelle Zusammenstellung der Titel für eine Musik CD. Drittens, die inhaltliche Konfiguration der Aggregationstiefe der einzelnen Informationsgüter im konfigurierten Makrobündel. Beispielsweise kann die Darstellung der Nachrichten auf der Startseite eines Nachrichtenportals entsprechend der Präferenzen eines Nutzers in der Form angepasst werden, dass zum einen nur solche Nachrichten überhaupt angezeigt werden, für die sich der Nutzer in der Vergangenheit interessiert hat. Zum anderen werden innerhalb der für ihn interessanten Nachrichten-kategorien nur diejenigen Überschriften mit einem Begleittext versehen, die für den Nutzer die höchste Relevanz haben, die übrigen Nachrichten dagegen nur als Überschriften in Linkform aufgelistet.

¹¹² Vgl. Rawolle (2002), S.16

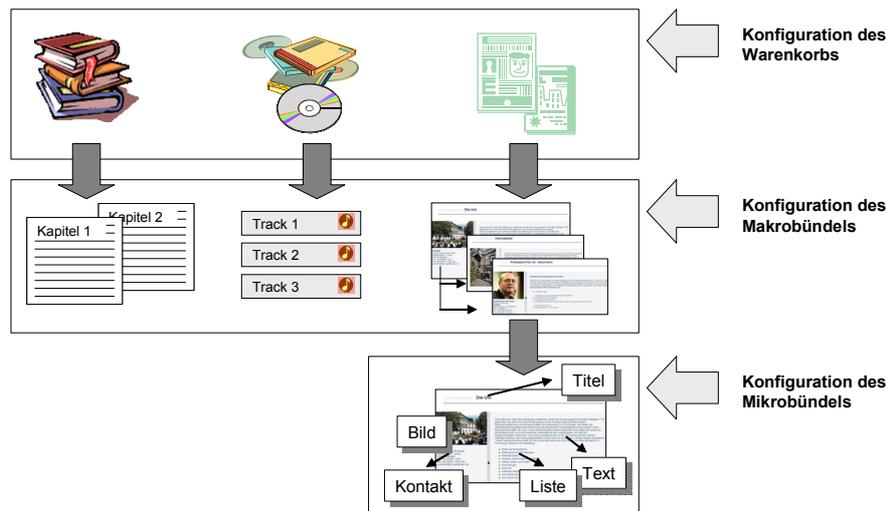


Abbildung 25: Individualisierungsmöglichkeiten für Medienprodukte

Vor dem Hintergrund der derzeit bereits in der Praxis beginnenden Erprobung sämtlicher Instrumente zur Individualisierung von Diensten und Dienstpräsentationen im Internet, liegt die Herausforderung für eine im Idealfall automatische Konfiguration von digitalen Produkten in Bezug auf individuelle Nutzerpräferenzen in einer, entsprechend der gewählten Filtermetrik, detaillierten Granularität der Merkmalsmodellierung der digitalen Produktressourcen für die Anwendung des Filterverfahrens. Hier besteht zukünftig noch weiterer Klärungsbedarf.

5 Zusammenfassung

Im zweiten Kapitel wurde ein Überblick gegeben über die verschiedenen Ursachen und Ausprägungsformen der gewachsenen Bedeutung des Individuums. Im dritten Kapitel wurden Möglichkeiten vorgestellt, wie die Bedürfnisse eines Individuums nach derzeitigem Stand von Wissenschaft und Technik identifiziert werden können. Dabei wurde insbesondere zwischen solchen Verfahren unterschieden, die auf einer expliziten Preisgabe des Nutzers beruhen, und solchen, die aus den protokollierten Aktivitäten verschiedener Nutzer typische Verhaltensmuster identifizieren und daraus Schlussfolgerungen über die impliziten individuellen Bedürfnisse ableiten. Im vierten Abschnitt wurden Systemkonzepte vorgestellt, die anhand identifizierbarer individueller Merkmale eines Nutzers eine automatische Individualisierung des Angebots im Internet erlauben. Eine Gegenüberstellung von eigenschaftsbasiertem Filtern und Collaborative Filtering einerseits in Verbindung mit dem Vergleich zwischen expliziten und impliziten Identifikationsverfahren andererseits deckte auf, dass nur eine Kombination zwischen den verschiedenen Verfahrenstypen eine sowohl hinsichtlich der allgemeinen Persönlichkeitsmerkmalen als auch hinsichtlich der situativen Bedürfnissen eines Nutzers lückenloses individuelles Angebot ermöglicht. Wie jedoch eine systematische Betrachtung der derzeit in der Praxis verfügbaren Individualisierungssysteme zeigt, entsteht insbesondere im Bereich einer automatisierten individuellen Konfiguration von digitalen Produkten weiterer Klärungsbedarf.

Literaturverzeichnis

[2003-03-31]

- Agrawal, Rakesh; Imielinski, Tomasz; Swami, Arun (1993): Mining Association Rules between Sets of Items in Large Databases; In: SIGMOD Record 21 (1993) Nr. 5; S. 207-216
- Amberg, Michael; Wehmann, Jens (2001): Situationsabhängige Dienste-Grundlagen ihrer Entwicklung; <http://www.wi3.uni-erlangen.de/forschung/situation/SAD.pdf> [2003-09-30]
- Bea, Franz Xaver; Haas, Jürgen (1995): Strategisches Management; Gustav Fischer Verlag, Stuttgart 1995
- Bensberg, Frank; Weiß, Thorsten (1999): Web Log Mining als Marktforschungsinstrument für das World Wide Web; In: Wirtschaftsinformatik 41 (1999) 5, S. 426-432
- Berger, Peter (1997): Individualisierung und sozialstrukturelle Dynamik; In: Beck, Ulrich; Sopp, Peter (Hrsg.): Individualisierung und Integration – Neue Konfliktlinien und neuer Integrationsmodus?; Leske+Budrich, Opladen 1997; S. 81-98
- Braybrooke, David; Lindblom, Charles (1965): A Strategy of Decision: Policy Evaluation as a Social Process; Free Press, New York 1965
- Cranor, Lorrie; Langheinrich, Marc; Marchiori, Massimo; Presler-Marshall, Martin; Reagle, Joseph (2002)
- Cranor, Lorrie; Weitzner, Daniel (2002): Summary Report - W3C Workshop on the Future of P3P; elektronisch veröffentlicht: <http://www.w3.org/2002/12/18-p3p-workshop-report> [2003-09-30]
- Davenport, Thomas; Prusak, Laurence (2000): Working Knowledge: how organizations manage what they know; Harvard Business School Press, Boston 2000
- EcIn (2002): Gute Perspektive für Paid Content; <http://www.ecin.de/news/2002/12/16/05190/>
- EcIn (2003): Steigende Akzeptanz für kostenpflichtige Inhalte; <http://www.ecin.de/news/2003/02/04/05336/> [2003-03-31]
- EU (1999): Grünbuch zur Konvergenz der Branchen Telekommunikation, Medien und Informationstechnologie und ihren ordnungspolitischen Auswirkungen; <http://europa.eu.int/scadplus/leg/de/lvb/l24165.htm> [2003-03-31]
- Fayyad, Usama; Piatetsky-Shapiro, Gregory; Smyth, Padhraic (1996): From Data Mining to Knowledge Discovery in Databases; elektronisch veröffentlicht: <http://www.kdnuggets.com/gpspubs/aimag-kdd-overview-1996-Fayyad.pdf> [2003-09-30]
- Frawley, William; Piatetsky-Shapiro, Gregory; Matheus, Christopher (1991): Knowledge Discovery in Databases: An Overview; In: AAAI Press, pp. 1-27, 1991
- Habermas, Jürgen (1999): Theorie des kommunikativen Handelns; Bd. 2: Kritik der funktionalistischen Vernunft; Suhrkamp Verlag, Frankfurt am Main 1999
- Haertsch, Patrick (2000): Wettbewerbsstrategien für die Digital Economy – Eine kritische Überprüfung klassischer Strategiekonzepte; Josef Eul Verlag, Köln 2000

- Hass, Berthold H. (2002): Geschäftsmodelle von Medienunternehmen – Ökonomische Grundlagen und Veränderungen durch neue Informations- und Kommunikationstechnik; Gabler, Wiesbaden 2002
- Hess, Thomas (2002): Implikationen des Internet für die Medienbranche – eine strukturelle Analyse; In: Keuper, Frank (Hrsg.): Electronic Business und Mobile Business. Ansätze, Konzepte und Geschäftsmodelle; Gabler, Wiesbaden 2002; S. 569-602
- Hess, Thomas (2002): Implikationen des Internet für die Medienbranche - eine strukturelle Analyse, In: Keuper, F. (Hrsg.): Electronic Business und Mobile Business - Ansätze, Konzepte und Geschäftsmodelle, Wiesbaden, S. 569 - 602, 2002.
- Keuper, Frank (2002): Convergence-based View – ein strategie-strukturierungstheoretischer Ansatz zum Management der Konvergenz digitaler Erlebniswelten; In: Keuper, Frank (Hrsg.): Electronic Business und Mobile Business. Ansätze, Konzepte und Geschäftsmodelle; Gabler, Wiesbaden 2002; S. 603-654
- Kirsch, Werner (1997): Strategisches Management – Die geplante Evolution von Unternehmen; Verlag Barbara Kirsch, München 1997
- Kirsch, Werner (1998): Betriebswirtschaftslehre – Eine Annäherung aus der Perspektive der Unternehmensführung; Verlag Barbara Kirsch, München 1998
- Kristol, David; Montulli, Lou (1997): HTTP State Management Mechanism; elektronisch veröffentlicht: <http://www.cis.ohio-state.edu/cgi-bin/rfc/rfc2109.html> [2003-09-30]
- Kübel, Rolf (1990): Ressource Mensch – Erfolg durch Individualität; C.H. Beck; München 1990
- Meffert, Heribert (2000): Marketing – Grundlagen marktorientierter Unternehmensführung; Gabler Verlag, Wiesbaden 2000
- Meier, Bernd (2003): Individualisierung - Ausführliche Begriffsbeschreibung; elektronisch veröffentlicht: http://www.chancenfueralle.de/Service___Termine/Lexikon//Individualisierung.html [2003-09-30]
- Melville, Prem; Mooney, Raymond; Nagarajan, Ramadass (2001): Content-Boosted Collaborative Filtering; elektronisch veröffentlicht: <http://www.cs.utexas.edu/users/ml/papers/cbcf-sigir-wkshp-01.pdf> [2003-09-30]
- Mertens, Peter (1994): Datenmustererkennung in der Ergebnisrechnung mit Hilfe der Cluseranalyse; In: DBW, 54. Jg. 1994, Heft 6, S. 739-753
- Merz, Michael (2000): Electronic Commerce – Marktmodelle, Anwendungen und Technologien; dpunkt-Verlag, Heidelberg 2000
- Miller, George (1956): The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information; The Psychological Review (1956), Nr. 63, S. 81-97; elektronisch veröffentlicht: <http://www.well.com/user/smalin/miller.html> [2003-09-30]
- Mings, Susan M.; White, Peter B (2000): Profiting from Online News: The Search for Viable Business Models; In: Kahin, Brian; Varian, Hal R. (Hrsg.): Internet Publishing and Beyond – The Economics of Digital Information and Intellectual Property; The MIT Press, Cambridge 2000; S. 62-96
- Mobasher, Bamshad (2002): WebPersonalizer: A Server-Side Recommender System Based on Web Usage Mining; elektronisch veröffentlicht: <http://facweb.cs.depaul.edu/research/TechReports/TR01-004.pdf> [2003-09-30]

- Mobasher, Bamshad; Cooley, Robert; Srivastava, Jaiveep (2000): Automativ Personalization based on Web usage Mining; elektronisch veröffentlicht: <http://maya.cs.depaul.edu/~classes/ect584/papers/mobasher.pdf> [2003-09-30]
- Mobasher, Bamshad; Dai, Honhua; Luo, Tao; Nakagawa, Miki (2001): Discovery and Evaluation of Aggregate Usage Profiles for Web Personalization; elektronisch veröffentlicht: <http://robotics.stanford.edu/users/ronnyk/WEBKDD-DMKD/Mobasher.pdf> [2003-09-30]
- Picot, Arnold; Reichwald, Ralf; Wigand, Rolf (1998): Die grenzenlose Unternehmung; Gabler, Wiesbaden 1998
- Piller, Frank Thomas (2001): Mass Customization – Ein wettbewerbsstrategisches Konzept im Informationszeitalter; Deutscher Universitäts-Verlag, 2001
- Porter, Michael (1999): Wettbewerbsstrategie; Campus Verlag, Frankfurt 1999
- Porter, Michael (2001): Strategy and the Internet (HBR OnPoint Enhanced Edition); Harvard Business School Publishing; Online Product Nr. 6358
- Runte, Matthias (2000): Personalisierung im Internet – Individualisierte Angebote mit Collaborative Filtering; Deutscher Universitäts-Verlag, Wiesbaden 2000
- Sabine Schockert (2003): Datenspuren im Internet; elektronisch veröffentlicht: <http://alice.zfn.uni-bremen.de/~a02h/spuren/logfile-auswertung.htm> [2003-09-30]
- Salton, G. ;Wong, A. ;Yang, C. (1975): A Vector Space Model for Automatic Indexing; In: Communications of the ACM, Nr. 18, S. 613-620, 1975
- Schumann, Jochen; Meyer, Ulrich; Ströbele, Wolfgang (1999): Grundzüge der mikroökonomischen Theorie; Springer, Berlin 1999
- Schweikart, Rudolf (1994): Organisation, Individualität, Konsum; LIT Verlag, Münster 1994
- Simon, Herbert Alexander (1957): Administrative Behavior; Free Press, New York 1957
- Shapiro, Carl; Varian, Hal R. (1999): Information Rules – A strategic guide to the network economy; Harvard Business School Press, Boston 1999
- Sonntag, Michael (1999): Das verborgene des Herzens – Zur Geschichte der Individualität; Rowohlt Verlag, Hamburg 1999
- Spiliopoulou, M.: Web Usage Mining: Data Mining über die Nutzung des Web. In: Hippner, H.; Küsters, U.; Meyer, M.; Wilde, K.D. (Hrsg.): Handbuch Data Mining im Marketing, Wiesbaden 2001, S. 489-510
- Staab, Steffen (2002): Wissensmanagement mit Ontologien und Metadaten; erschienen in: "Informatik_Spektrum", 20. Juni 2002-08-27
- Thielmann, Bodo (2000): Strategisches Innovationsmanagement in konvergierenden Märkten: Medien und Telekommunikationsunternehmen in Online-Diensten un im digitalen Fernsehen; Gabler, Wiesbaden 2000
- van Meteren, Robin; van Someren, Maarten (2000): Using Content-Based Filtering for Recommendation; elektronisch veröffentlicht: www.ics.forth.gr/~potamias/mlnia/paper_6.pdf [2003-09-30]
- Wittgenstein, Ludwig (1984): Vorlesungen 1930–1935; Suhrkamp Verlag, Frankfurt am Main 1984

- Wohlrab-Sahr, Monika (1997): Individualisierung: Differenzierungsprozess und Zurechnungsmodus; In: Beck, Ulrich; Sopp, Peter (Hrsg.): Individualisierung und Integration – Neue Konfliktlinien und neuer Integrationsmodus?; Leske+Budrich, Opladen 1997; S. 23-36
- Wu, Kun-Lung; Yu, Philip; Ballmann, Allen (1998): SpeedTracer: A Web Usage Mining and Analysis Tool. IBM Systems Journal 37(1): 89-105(1998)
- Zerdick, Axel; Picot, Arnold; Schrape, Klaus; Artopé Alexander; Gold-hammer, Klaus; Heger, Dominik; Lange, Ulrich; Vierkant, Eckart; López-Escobar, Esteban; Silverstone, Roger (2001): Die Internet-Ökonomie – Strategien für die digitale Wirtschaft; Springer-Verlag, Berlin/Heidelberg 2001