

WIRTSCHAFTSUNIVERSITÄT WIEN

DIPLOMARBEIT

Titel der Diplomarbeit:

Wirtschaftliche Bedeutung von Semantic Web Mining und Recommender Systemen

Verfasserin/Verfasser: Jasna Tusek

Matrikel-Nr.: 9950128

Studienrichtung: Betriebswirtschaft

Beurteilerin/Beurteiler: o.Univ.Prof. Dkfm.Dr.rer.comm. Wolfgang Janko

Ich versichere:

dass ich die Diplomarbeit selbstständig verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt und mich auch sonst keiner unerlaubten Hilfe bedient habe.

dass ich dieses Diplomarbeitsthema bisher weder im In- noch im Ausland (einer Beurteilerin/ einem Beurteiler zur Begutachtung) in irgendeiner Form als Prüfungsarbeit vorgelegt habe.

dass diese Arbeit mit der vom Begutachter beurteilten Arbeit übereinstimmt.

Datum

Unterschrift

Für meine Eltern

Wirtschaftliche Bedeutung von Semantic Web Mining und Recommender Systemen

Economic Relevance of Semantic Web Mining and Recommender Systems

Stichworte: Customer Relationship Management, Data Mining, Individualisierung, Ontologie, Personalisierung, Empfehlungssysteme, Semantic Web Mining

Keywords: Customer Relationship Management, Data Mining, Individualization, Ontology, Personalization, Recommender Systems, Semantic Web Mining

Zusammenfassung

Der zunehmende Überfluss an Informationen im weltweiten Netzwerk macht den Informationsaustausch immer schwieriger. Die Suchmaschinen werden zwar zur Informationsfilterung eingesetzt, können aber keinen erfolgreichen Informationsaustausch garantieren. Denn Informationen können mehrere Bedeutungen haben und von Suchmaschinen wegen fehlender Semantik nicht richtig interpretiert werden. Die Web-Technologie der ersten Web-Generation erlaubt die Darstellung von Informationen lediglich in maschinen-lesbaren Form. Das Web der zweiten Generation, das semantische Web, erlaubt zudem die Darstellung in maschinen-interpretierbarer Form, wodurch die Interaktion zwischen Mensch und Maschine verbessert wird. Die Technologie des semantischen Webs kann daher die wirtschaftlichen Aktivitäten im Web fördern, indem beispielsweise die Suchanfragen optimiert werden. Die in diesem Zusammenhang zur Anwendung kommenden Instrumente der Informationstechnologie, insbesondere *Semantic Web Mining* und *Recommender Systeme*, sind bedeutende Gegenstände dieser Arbeit.

Die wirtschaftliche Bedeutung der in dieser Arbeit präsentierten Web-Technologie wird auf Basis der identifizierten Einflussfaktoren erarbeitet. Im Vordergrund steht die Idee der Verbesserung des Unternehmenserfolgs durch Ausschöpfung des wirtschaftswissenschaftlichen und technologischen Potentials.

Die Arbeit zeigt abschließend die wirtschaftliche Bedeutung der informationstechnologischen Instrumente für die Praxis. Ebenso werden die bedeutenden Ziele und Potentiale der Einflussfaktoren aufgezeigt.

Abstract

The increasing information overload of the World Wide Web tends to hamper the information exchange. Search engines are used for information filtering, but do not guarantee the successful exchange of information. Due to different meanings in information and missing semantics the search engines are not always able to process information correctly. The technology of the first web generation allows the presentation of information merely in machine-readable way. The Web of the next generation, the

Semantic Web, allows furthermore the presentation in machine-understandable way, whereby the interaction between human and computer is enhanced. So the technology of Semantic Web facilitates the economic activities on the Web by optimizing queries for instance. In this context some special techniques of information business are applied, in particular the Semantic Web Mining and Recommender Systems and are therefore important topics of this thesis.

The economic relevance of Web technology, especially of Semantic Web Mining and Recommender Systems, will be presented on the basis of the identified *influencing factors*. The idea of improving the economic success by utilization of applied economic and technological potential comes to the fore.

Finally, the thesis treats the economic relevance of information technology for practice. Also the significant aims of identified factors and the utility will be presented.

Kernpunkte für das Management

Diese Arbeit basiert auf der grundlegenden Überlegung der identifizierten Einflussfaktoren für das Web. Besonderes Augenmerk wird auf die Präsentation der Einflussfaktoren und der dazugehörigen Instrumente, ihre Relevanz für die Praxis und auf die wirtschaftliche Bedeutung der in diesem Zusammenhang präsentierten Instrumente, *Semantic Web Mining* und *Recommender Systeme*, gelegt. Anhand der präsentierten Theorie der Einflussfaktoren werden folgende Punkte erarbeitet:

- Die Signifikanz des Semantic Web Mining sowie der relevante Einsatz von Empfehlungssystemen in der Praxis werden anhand der identifizierten Einflussfaktoren aufgezeigt.
- Im analytischen Teil dieser Arbeit werden Unternehmen aus dem Bereich des Mobilfunknetzes präsentiert und anhand der in Praxis häufig zur Anwendung kommenden Unternehmensphilosophie des *Customer Relationship Managements* die Relevanz der Einflussfaktoren für die Praxis dargestellt.
- Die identifizierten Einflussfaktoren für das Web stellen das Grundkonzept des vorliegenden Themas dar und die anschließende Erarbeitung der praxisorientierten Analyse dient der besseren Nachvollziehbarkeit der Ergebnisse der Arbeit.

INHALTSVERZEICHNIS

1 EINFÜHRUNG	3
2 PROBLEMSTELLUNG UND LÖSUNGSANSÄTZE	6
2.1 PROBLEMSTELLUNG	6
2.1.1 Informationsüberfluss.....	7
2.2 LÖSUNGSANSÄTZE FÜR KUNDEN	7
2.2.1 Individualisierung	8
2.2.2 Personalisierung	9
2.3 LÖSUNGSANSÄTZE FÜR UNTERNEHMEN.....	10
2.3.1 Einheitliche Standards für Informationsaustausch im Web	10
2.3.2 Technologie mit maschinen-verarbeitbarer Semantik.....	11
3 EINFLUSSFAKTOREN	13
3.1 TECHNOLOGIE FÜR DAS WEB	13
3.1.1 Semantic Web	16
3.1.2 Ontologien.....	22
3.2 INFORMATION AUS DEM WEB.....	30
3.2.1 Data Mining	30
3.2.2 Semantic Web Mining.....	38
3.3 PERSONALISIERUNG IM WEB.....	42
3.3.1 CRM vs. eCRM.....	42
3.3.2 Recommender Systeme	45
4 ANALYSE	52
4.1 ZIELE DER EINFLUSSFAKTOREN	53
4.2 SIGNIFIKANZ DER EINFLUSSFAKTOREN FÜR DIE PRAXIS.....	55
4.2.1 Das Mobilfunknetz.....	56
4.2.2 Optimierung der Web-Dienste.....	61
4.2.3 Optimierung der kommerziellen Webseiten.....	69
4.3 SIGNIFIKANZ DER EINFLUSSFAKTOREN FÜR ANWENDER.....	73
4.3.1 Semantic Web	74
4.4 SIGNIFIKANZ DER EINFLUSSFAKTOREN FÜR UNTERNEHMEN	75
4.4.1 Semantic Web Mining.....	76
4.4.2 Recommender Systeme	77
5 FAZIT	80

LITERATURVERZEICHNIS	81
ABKÜRZUNGSVERZEICHNIS	95
BILDVERZEICHNIS	97
TABELLENVERZEICHNIS	97
ANHANG	98
1 DIE VOLLSTÄNDIGE WSDL 1.2 SYNTAX.....	98

1 EINFÜHRUNG

„The Semantic Web and Web Services are envisioned as the enabling technologies for the next generation of web applications.” [DoRS05]

Das weltweite Netzwerk (Web) besteht ausschließlich aus *Informationen* sei es in Form von Texten, Graphiken oder Animationen. Diese Informationen werden auf unzähligen kommerziellen, privaten oder öffentlichen Webseiten präsentiert. Die dabei zur Anwendung kommende Technologie macht die Informationsrepräsentation im Web erst möglich. Das *World Wide Web* stellt somit den Zugang zu Informationen aus der ganzen Welt dar und bietet Möglichkeiten zur Informationsfilterung. Aufgrund der unzähligen Webseiten entsteht jedoch das Problem des Informationsüberflusses im Web. Dadurch wird die Filterung von relevanten Informationen im Web erschwert. Die Web-Technologie stellt aber die wesentliche Grundlage für die Verbesserung des Informationsaustausches, als Beispiel kann hierbei das semantische Web genannt werden. Das *Semantic Web* erlaubt eine beachtenswerte Verbesserung im Bereich der Informationsfilterung und des Informationsaustausches, indem man der Information eine entsprechend definierte Bedeutung zuordnet. Durch festgelegte Definitionen von Begriffen können Anfragen, welche zum Beispiel an Suchmaschinen gestellt werden, genauer beantwortet werden. Im semantischen Web können Suchmaschinen die Information besser *verstehen* und sie daher auch exakter bearbeiten. Aus der technologischen Perspektive ist das semantische Web eine Verbesserung des aktuellen Webs im Sinne einer technologischen Erweiterung. Somit stellt das semantische Web eine Innovation der Web-Technologie dar, welche insbesondere durch wirtschaftliche Umsetzung dieser Technologie realisiert wird. Im Zusammenhang mit wirtschaftlichen Aktivitäten stellt jede Innovation einen wesentlichen Einflussfaktor dar. In der vorliegenden Arbeit wird auf die Innovation der Informationstechnologie eingegangen, das heißt es werden insbesondere wirtschaftliche Aktivitäten im Web betrachtet und die dabei zur Anwendung kommenden informationstechnologischen Instrumente.

Auf der Grundlage von technologischen Innovationen können neue Instrumente zum Einsatz kommen. Das naheliegendste Beispiel im Zusammenhang mit dem vorliegenden Thema ist das *Semantic Web* und das damit realisierbare *Semantic Web Mining*, welche in Kapitel 3 erläutert werden. Das semantische Web ermöglicht die Entwicklung und den Einsatz neuer Instrumente, das heißt neue technologische Mittel können zur Anwendung kommen. Durch Innovationen und Verbesserung der bestehenden Web-Technologie ist die Realisierung von neuen *Technologien für das Web* erst möglich. Diese Technologie übt Einfluss auf wirtschaftliche Aktivitäten im Web und somit auch auf den Erfolg des Unternehmens aus. Zweckmäßig wird daher der erste identifizierte Einflussfaktor „Technologie für das Web“ genannt.

Der Einsatz dieser neu gewonnen Technologien bzw. neuen informationstechnologischen Instrumenten bietet die Grundlage für detaillierte Beschaffung von „Information aus dem Web“, welche als zweiter Einflussfaktor identifiziert wird. Die aus dem Web gewonnen Informationen üben Einfluss auf das Anwendungsgebiet. Das Unternehmen kann die relevanten Informationen über Web-Besucher auswerten und analysieren. Das *Semantic Web Mining* erlaubt durch maschinen-verarbeitbare Semantik genauere Informationsgenerierung im Web. Das Ziel des semantischen Webs, welches die Grundlage für das

Semantic Web Mining darstellt, ist es verbesserte Kommunikation zwischen Mensch und Maschine zu gewährleisten.

Die *wirtschaftliche Bedeutung* des Webs und der Web-Technologie spiegelt sich insbesondere im dritten Einflussfaktor durch die „Personalisierung im Web“ wider. Durch die dabei zur Anwendung kommende Technologie lassen sich Kunden übers Web individuell ansprechen, sowie die Zufriedenheit der Stammkunden steigern. Das ebenfalls nahe liegende Beispiel für eine solche Technologie ist das Recommender System. Die grundlegende Überlegung für die vorliegende Diplomarbeit stellen daher folgende Einflussfaktoren für das Web dar:

- *Technologie für das Web*
- *Information aus dem Web*
- *Personalisierung im Web*

Der Zusammenhang und die Bedeutung dieser Einflussfaktoren für die Wirtschaft werden anhand der folgenden Überlegung dargestellt. Ein hauptsächlich online-tätiges Unternehmen ist daran interessiert, relevante Informationen über seine Kunden aus dem Web zu filtern, durch seinen Web-Auftritt die Kundenbeziehung zu pflegen bzw. neue Kunden zu akquirieren und sein Web-Potential durch die Verbesserung der Technologie und durch Innovationen auszuschöpfen. Genauso gut lässt sich argumentieren, dass die Realisierung von *Technologien für das Web* durch technologischen Fortschritt zur effizienteren Beschaffung von *Informationen aus dem Web* führt. Die dabei gewonnenen Informationen über Kunden können gesammelt werden und zum Zwecke der Pflege von Kundenbeziehung durch Personalisierungssoftware wie zum Beispiel der Recommender Systeme analysiert werden. Durch die Anwendung von Personalisierungssoftware im Web lassen sich die Ziele dieser *Personalisierung* wie zum Beispiel die Intensivierung der Kundenbeziehung und Kundengewinnung realisieren. Da die Personalisierung im Web den Zielen der Unternehmensphilosophie des Customer Relationship Managements (CRM) dient, spricht CRM ebenfalls für den Einsatz von Personalisierungssoftware. Im Informationszeitalter ist die innovative Web-Technologie der primäre Einflussfaktor für die erfolgreiche Realisierung der Kundenbeziehung im Web.

Die genannten Einflussfaktoren für das Web unterscheiden sich von den üblichen Einflussfaktoren dadurch, dass sie insbesondere mithilfe von Web erst realisiert werden können. Übliche Faktoren wie Marketing, gute Organisations- und Investitionsentscheidungen sowie beschaffungs- und produktionsseitige Anforderungen, die für den betrieblichen Erfolg ebenfalls ausschlaggebend sind, werden hierbei vorausgesetzt und bedürfen daher keiner Details.

Im Zusammenhang mit der Geschwindigkeit der Geschäftsprozesse, welche durch die innovative Technologie gewährleistet wird, lässt sich das Zeitmanagement gut planen, da durch rasche Geschäftsprozessabwicklungen viel Zeit sowohl seitens des Unternehmens als auch seitens der Kunden gespart wird. Die Zeiteinsparung lässt sich durch das semantische Web realisieren und durch die Anwendung der informationstechnologischen Instrumente, welche unter anderem in dieser Arbeit präsentiert werden, ausschöpfen.

Von ausschlaggebender Bedeutung ist hierbei die gleichwertige Betrachtungsweise zwischen Unternehmen und Kunden: der Erfolg des Unternehmens spiegelt sich im Erfolg des Kunden wider und vice versa. Ist der Kunde mit seiner Anschaffung zufrieden, wird er auch in Zukunft zum ökonomischen Erfolg des Unternehmens beitragen. Dies lässt sich sehr gut anhand der Unternehmen des

Mobilfunknetzes gut erkennen, welche daher für den praxisorientierten Teil (Kapitel 4) herangezogen werden.

Der Erfolg des Unternehmens spiegelt sich jedoch vor allem in der Zufriedenheit des Kunden wieder unabhängig von der Art der Produkt-Anschaffung. Wird das Produkt in einer Filiale angeschafft, fallen Versandkosten weg – bei Online-Bestellungen dagegen spart der Kunde Zeit und Nerven. In diesem Zusammenhang lässt sich auch vom Erfolg des Kunden sprechen, denn jeder Kunde hat einen erfolgreichen Kauf getätigt, wenn er mit dem Produkt oder der Leistung zufrieden ist. Aus diesem Grund wird insbesondere dem *Customer Relationship Management* wesentliche wirtschaftliche Bedeutung zugeordnet und daher im Zusammenhang mit der *Personalisierung im Web* präsentiert.

In Kapitel 2 werden die Problemstellung und die Lösungsansätze diskutiert. Anschließend werden in Kapitel 3 die drei genannten Einflussfaktoren detaillierter beschrieben und die dabei jeweils zur Anwendung kommenden informationstechnologischen Instrumente präsentiert. In Kapitel 4 werden die Einflussfaktoren und die informationstechnologischen Instrumente hinsichtlich ihrer wirtschaftlichen Bedeutung für die Praxis untersucht. Zudem werden auch die Ziele der Einflussfaktoren anschaulich dargestellt. Somit ist die Arbeit grundlegend in einen theoretischen Teil (Kapitel 2 und 3) und in einen praxisorientierten Teil (Kapitel 4) gegliedert.

2 PROBLEMSTELLUNG UND LÖSUNGSANSÄTZE

Das Web nimmt bei der effizienteren Erschließung von Arbeitsabläufen sowie bei Erschließung neuer Märkte zentrale Rolle ein [HaSZ02]. Zudem bringt der Trend der weit verbreiteten und zunehmenden Anwendung des Webs insbesondere mit dem technologischen Fortschritt im Bereich der Hardware und Software bedeutende Nutzenpotentiale mit sich.

Mit *Electronic Commerce (E-Commerce)* wird in dieser Arbeit der Bereich des *Business-to-Consumer (B2C)* spezifiziert. Laut Fensel et al. [FHLW03] ist E-Commerce vor allem aus zwei Gründen ein wichtiger und wachsender Wirtschaftszweig:

1. Durch Erweiterung vorhandener Wirtschaftsmodelle werden Kosten reduziert und existierende Vertriebskanäle realisiert bzw. neue Vertriebsmöglichkeiten hinzugefügt. Ein Beispiel solcher Erweiterung des Wirtschaftsmodells sind *Online Shops*.
2. Die Entstehung komplett neuer Wirtschaftsmodelle wird ermöglicht bzw. vorhandene Modelle erlangen durch E-commerce größere Bedeutung. Beispiele für neue Wirtschaftszweige, die vom *Electronic Commerce* resultieren, sind *Shopping Agents*, *Online Market Places* und *Auction Houses*. Letzteres vergleicht Shopping Prozesse bzw. Einkäufe mit eigenem, signifikantem Einnahmefluss. Die Vorteile des *Online Shops* und der dadurch erzielte Erfolg führten zur rasanten Vermehrung solcher Shopping Webseiten bzw. zum vermehrten Einsatz dieser Praktika.

Die Aufgabe des neuen Web-Konsumenten ist nun einen Online Shop zu finden, das das vom Web-Konsumenten nachgefragte Produkt zum gewünschten Zeitpunkt in der gewünschten Qualität und Quantität sowie zum fairen Preis anbietet [FHLW03]. Dieses Ziel via Browsing zu erreichen erfordert immensen Zeitaufwand, sowie eine gründliche und detaillierte Recherche, deren Intensität abhängig von Erfahrungswerten von Web-Konsument zu Web-Konsument schwankt.

In folgenden Subkapiteln wird auf das Problem des Informationsüberflusses und auf die Lösungsansätze dieses Problems näher eingegangen. Dabei wird zwischen Anbieter und Anwender der Informationstechnologie unterschieden. Die Unternehmen sind die Anbieter und die Kunden zählen zu den Anwendern der Informationstechnologie. Daher wird auch in weiterer Folge zwischen Lösungsansätzen für Unternehmen und Lösungsansätzen für Kunden differenziert.

2.1 Problemstellung

Die Bedeutung des Informationszeitalters liegt insbesondere in der Forschung und Entwicklung der IT-basierten Anwendungen, welche wiederum aufgrund ihrer nutzenstiftenden Eigenschaft auch immer häufiger zur Anwendung kommen. In Verbindung mit der notwendigen Hardware lassen sich die

Anwendungen des Webs für optimierte Terminplanung einsetzen und die damit verbundene Flexibilität ausschöpfen. Durch das reichlich zur Verfügung stehende Informationsangebot im Web wird das Problem der Unüberschaubarkeit der vorhandenen Informationen identifiziert. Aufgrund dieses Problems wird die Optimierung der ökonomischen Terminplanung und des gesamten Zeitmanagements begrenzt.

2.1.1 Informationsüberfluss

Als interaktives Medium dient das Web hauptsächlich der Informationsverbreitung. Das facettenreiche und dynamische Medium ist durch dessen Beliebtheit enorm gewachsen, was zur der Situation des Informationsüberflusses geführt hat [Maes94].

Der Informationsüberfluss im Web stellt an Kunden enorme Herausforderungen bei der Selektierung von Online-Produkten dar. Diese Herausforderung entsteht auch für die Unternehmen, die hauptsächlich online agieren in dem Versuch die Präferenzen der Konsumenten effizient zu identifizieren. Verschiedene Empfehlungssysteme mit der unterschiedlichen Repräsentation von Daten und Empfehlungsmethoden werden als Lösungsansatz für diese Herausforderungen herangezogen [ZaWH04].

Zan et al. [ZaWH04] entwickeln ein Modell, das die Daten-Repräsentation generisch zur Verfügung stellt und das verschiedene Empfehlungsmodelle unterstützen kann. Um die Nützlichkeit und Flexibilität dieses Modells aufzuzeigen, werden folgende drei Empfehlungsmethoden entwickelt: *Direct Retrieval*, *Association Mining* und *High-degree Association Retrieval* [vgl. ZaWH04].

Das Web stellt enorme Informationsquellen fast jeglicher Art zur Verfügung, das sich von DNA Databases bis zu Menüs verschiedener Restaurants erstreckt. Dennoch sind diese Informationen unter verschiedenen Web Servern und Hosts dermaßen verstreut, so dass sie in Unüberschaubarkeit und Informationsüberfluss münden. Zudem verwenden die verschiedenen Web Server und Hosts wiederum unterschiedliche Formate. Wenn diese Informationseinheiten vom Web extrahiert werden könnten um in eine strukturierte Form integriert zu werden, würden diese strukturierten Informationseinheiten eine noch nie da gewesene, insbesondere überschaubare Informationsquelle bieten [Brin03]. Das daraus resultierende Potential für die Wirtschaft aber auch für das Studium, Berufs- und Privatleben ließe sich somit ausschöpfen.

2.2 Lösungsansätze für Kunden

Der Informationsüberfluss stellt aufgrund der dabei auftretenden Unüberschaubarkeit in der Zeit der Informationstechnologie ein Defizit dar, woraufhin Lösungsansätze zur Anwendung kommen um das Potential des Informationszeitalters im Web besser zu realisieren. Die Menge und Art von Informationen lässt sich richtig einsetzen und das daraus resultierende Wissen in der Praxis gut umsetzen, wenn die Daten selektierbar, sortierbar und insbesondere auffindbar sind. Die Art und Weise wie nun der Inhalt einer Webseite präsentiert wird, entscheidet darüber ob die Informationen im Netz auffindbar sind und somit darüber ob ein zusätzlicher Käufer erworben wird. Zum anderen entscheidet die Präsentationsart des Inhalts auch darüber ob Kunden sich angesprochen fühlen und mit der Dienstleistung der Webseite zufrieden gestellt werden können. Die Untersuchung dieser Sachverhalte hat in der vorherrschenden

Literatur zu den Lösungsansätzen der Individualisierung und Personalisierung geführt. Die Praxis bestätigt den erfolgreichen Lösungsweg zunehmend: “[...] *personalization becomes a popular remedy to customize the Web environment for users.*” [ShCh03] Die anwenderspezifischen Dienstleistungen bieten dem Anwender den notwendigen Mehrwert und steigern generell die Zufriedenheit. Daher wird im Folgenden auf die Ansätze der Individualisierung und der Personalisierung eingegangen.

2.2 1 Individualisierung

Die Individualisierung stellt definitionsgemäß den Prozess der Erstellung eines *individuellen Gutes* dar. Wobei ein individuelles Gut ein Leistungsbündel ist, welches sich aus den Teilleistungen *Leistungsspezifikation*, *Leistungszuordnung* und *Leistungserbringung* zusammensetzt [Scha03]. Diese Definition des Begriffs wird in weiterer Folge ergänzt.

Die Individualisierung ist ein aus Teilprozessen bestehender Prozess, der aus dem *Inputfaktor* „Information“ neue Informationen als *Outputfaktor* generiert. Das Resultat der Prozesse der Individualisierung ist es, für den Kunden die nutzenmaximale Leistung in Form von personalisierter Information zu produzieren. Individualisierung stellt somit ein Entscheidungsproblem dar, in dem entschieden werden soll, welche Leistung dem Kunden den größten Nutzen stiftet [Scha03]. Jene Leistung bzw. jene interessante Information, die den Kunden anspricht und ihm daher auch den größten Nutzen stiftet, wird personalisiert. Welche Informationen nun personalisiert werden, hängt von individuellem Benutzer und seinen Interessen ab.

Die Individualisierung wird aus entscheidungstheoretischer Sicht laut der zugrunde liegenden Literatur folgendermaßen beschrieben: der aktive Benutzer sucht zur Lösung seines persönlichen Problems eine Leistung bzw. einen Entscheidungsraum. Das Kriterium für die Auswahl der Leistung stellt der *Nutzen* dar, nämlich jener Nutzen, den die Leistung dem Benutzer stiftet. Dieser Nutzen entspricht dem Maß, in welchem sich die Leistung zur Lösung des persönlichen Problems eignet [Scha03]. Der Individualisierungsgrad gibt an, in wie weit der Kunde in die Produktionsprozesse integriert ist bzw. wie gut das bestimmte Produkt die Bedürfnisse des Kunden zu befriedigen vermag [Scha03].

Eine sehr große Anzahl potenzieller Leistungen können identifiziert werden, welche horizontal differenziert sind. Somit ist der durch eine Leistung gestiftete Nutzen problem- und ebenso personenspezifisch [Scha03]. Im Folgenden wird ein anschauliches Praxisbeispiel präsentiert.

Beispiel:

Ein potentieller Kunde beabsichtigt für den privaten Gebrauch ein mobiles Telekommunikationsgerät zu erwerben. Da er sich sowohl Zeit als auch Benzinkosten sparen möchte, kommt für ihn nur der Erwerb über ein Online Shop in Frage. Zur Lösung dieses Problems benötigt er insbesondere für ihn relevante Informationen über das gewünschte Produkt. Dabei kann er verschiedene Web-Leistungen in Anspruch nehmen. So kann er beispielsweise mit Suchmaschinen im Web (*google.at*, *yahoo.com*, *etc.*) recherchieren oder direkt auf Webseiten der Netzbetreiber seines Landes. Beispiele solcher Webseiten in Österreich sind *a1.net*, *one.at*, *t-mobile.at* etc. Zudem kann er direkt auf Webseiten der Produzenten solcher Produkte recherchieren, wie zum Beispiel *nokia.com*, *motorola.com* oder *samsung.com*.

Jede dieser Leistungen stiftet dem potentiellen Web-Kunden unterschiedlichen Nutzen. Während die Suchmaschine Google auf die Suchabfrage lediglich die übereinstimmenden Ergebnisse zwischen dem gesuchten Wort und jenem in Dokument liefert, ist die Recherche auf spezifischen Webseiten der Netzbetreiber schon insofern eingegrenzt, dass die entsprechende Trefferquote gewährleistet ist. Denn diese Webseiten bieten ausschließlich Dienste um mobile Telekommunikationsgeräte an. (Die Webseiten der Netzbetreiber *al.net* und *one.at* werden in Kapitel 4.2.1 zum Thema.)

In diesem Fall sucht der potentielle Kunde ein Gerät mit bestimmten Funktionen. Da er das Gerät gerne auf Reisen mitnehmen möchte, stellt er auch hohe Anforderungen an die Funktionen des Gerätes. Einerseits soll das mobile Telekommunikationsgerät einige grundlegende Features aufweisen, wie zum Beispiel einen Kalender und einen Wecker. Andererseits erwartet er sich eine Kamera mit mindestens 1 Megapixel und vierfachem Digitalzoom, einen MP3-Player und Radio sowie zusätzliche Web-Dienste wie zum Beispiel das *Wireless Application Protocol* (WAP) oder i-mode.

Die eigenständige Recherche verursacht Kosten in Form von Internetverbindungskosten und Opportunitätskosten durch die mit der Recherche verbrauchten Zeit. In Abhängigkeit der Vorgangsweise bei der Recherche wird ein Web-Besucher an seine Informationen eventuell rascher gelangen als ein anderer. Die Art der Information ist für die Dauer der Recherche ebenfalls ausschlaggebend.

In der Praxis basiert die Recherche nach Informationen zum Großteil auf Anwendungen des Webs. Das Beispiel zeigt, dass die Beschaffung von *Informationen aus dem Web* in der Praxis eine wesentliche Komponente zur Anschaffung von gewünschten Informationen über Produkte darstellt. In weiterer Folge wird versucht die Recherche zu optimieren. Dazu reicht das Know-how des Anwenders nicht aus. Vielmehr ist die Technologie, welche die Recherche erst ermöglicht, zu optimieren und in weiterer Folge benutzerfreundlich sowie maschinenfreundlich zu gestalten. Auch aus diesem Grund wird in Kapitel 3 auf die neuen Ansätze des *Semantic Web* und *Semantic Web Minig* näher eingegangen.

2.2.2 Personalisierung

Die Personalisierung lässt sich in unterstützende und vollautomatisierte Personalisierung unterteilen. Ersteres trifft zu, wenn dem Personalisierungssystem eine Komplementfunktion zugeschrieben wird, und somit den Berater unterstützt. Die vollautomatisierte Personalisierung entspricht dem Prozess der Individualisierung, der autonom durch das Personalisierungssystem durchgeführt wird. Die Qualität der Personalisierungsleistung beschreibt das Ausmaß, mit welchem die Merkmale der Leistungsspezifikation in der Lage sind, das Kundenproblem zu lösen. Die Personalisierungsintensität ist ein Maß zur Beschreibung der Qualität einer individuellen Leistung auf Basis des Nutzens, den diese Leistung einem einzelnen Kunden stiftet. Alternativ zur Nutzenbetrachtung lässt sich zur Messung der Qualität das Maß der Kundenzufriedenheit anwenden. Die Kundenzufriedenheit misst die Differenz zwischen den Erwartungen des Kunden und der subjektiv wahrgenommenen Leistung [Scha03].

Wesentliche Faktoren der Personalisierung stellen zum einen die notwendigen Daten und zum anderen das Vertrauen des Kunden in den Anbieter. Für die Leistungsspezifikation sind detaillierte Daten,

Informationen und Wissen über den Kunden und dessen Problem notwendig. Das Vertrauen des Kunden in den Anbieter ist insbesondere dann relevant, wenn der Kunde die Qualität der Leistung nicht beurteilen kann [Scha03].

Zu einer der wesentlichen Personalisierungsstrategien gehört das Recommender System. Durch die Aufgabe des Recommender Systems wird der Endbenutzer in seinen Anwendungen unterstützt und beraten. Abhängig von den Präferenzen des Konsumenten werden vom System spezifische Produkte selektiert und empfohlen. Auf Basis dieser Empfehlungen kann der Endbenutzer rascher eine Entscheidung treffen. Die empfohlenen Angebote stellen eine übersichtliche Entscheidungsgrundlage für den Konsumenten dar. Vor allem in der Zeit des Informationsüberflusses ist jede Unterstützung für Konsumenten nutzenstiftend. Aus diesem Grund gehört das Recommender System zu einer der wesentlichen Technologien im *Electronic Business*, die mit dem Wachstum des Webs weiter an Bedeutung zunimmt.

Die Recommender Systeme erfüllen primär eine Beraterfunktion und dienen der Kundenzufriedenheit und der Steigerung der Kundenloyalität. Aufgrund des sinnvollen Einsatzes der Recommender Systeme wird in Kapitel 3.3.2 näher darauf eingegangen.

2.3 Lösungsansätze für Unternehmen

Im Zusammenhang mit wirtschaftlichen Aktivitäten führt der Wettbewerb zwischen Unternehmen zur Forderung nach Systemverbesserung und verbesserten Organisationsstrukturen. Das Erkennen von Verbesserungspotentialen und Realisierung dessen sorgt für bessere Qualität in der Leistung des Unternehmens, wodurch sich das Unternehmen am Markt gegenüber Konkurrenz erfolgreich durchsetzen kann.

2.3.1 Einheitliche Standards für Informationsaustausch im Web

Ursprünge im Austausch elektronischer Daten zwischen verschiedenen Gesellschaften finden sich bereits in Wirtschaftsprozessen der 60er Jahren. Informationen über wirtschaftliche Transaktionen auszutauschen erfordert einheitliche Standards, die sowohl vom Sender als auch vom Empfänger eingehalten werden müssen. Diese einheitlichen Standards beinhalten ein *Protokoll* für die Übermittlung des Inhalts und eine einheitliche Sprache zur Beschreibung des Inhalts. Diesem Zweck entstammte auch die Norm des *Electronic Data Interchange for Administration, Commerce, and Transport (EDIFACT)*, welche zur internationalen Übertragung von Daten herangezogen wird [FHLW03].

Im Laufe der Jahrhundertwende hat sich für die besondere Eigenschaft der kompatiblen Systeme untereinander der Begriff der *Interoperabilität* durchgesetzt. Definitionsgemäß wird unter Interoperabilität die Fähigkeit zur Zusammenarbeit von verschiedenen Systemen, Techniken oder Organisationen verstanden. Sind zwei Systeme miteinander kompatibel, werden gemeinsame Standards eingehalten. Somit können Informationen auf effiziente Weise ausgetauscht bzw. dem Benutzer zur Verfügung gestellt werden [Wiki05]. Ein Beispiel für eine Art der Interoperabilität ist das *World Wide Web* selbst, durch welches jede kompatible Software Daten verknüpfen und austauschen kann [Gate05].

Informationstechnologisch betrachtet steht Interoperabilität für verschiedene Programme, die dasselbe Dateiformat oder dieselben Protokolle verwenden können. Interoperabilität hat vor allem in der Wirtschaft eine bedeutende Funktion. Kompatible Programme bieten konkurrierenden Unternehmen die Chance auf dem Markt Fuß zu fassen. Inkompatible Programme dagegen sorgen für Monopolstellung am Markt. Für fairen und freien Wettbewerb ist es jedoch erforderlich Monopole aufzulösen [Wiki05].

Die Informationsgewinnung, -verarbeitung und -austauschbarkeit sind wesentliche Bereiche des Webs unabhängig vom Gesichtspunkt des Betrachters. Zum Beispiel benötigt ein Unternehmen Informationen über seine Kunden. Ein anderes Beispiel stellt ein potentieller Kunde dar, der einen Preisvergleich zwischen den Angeboten von mobilen Telekommunikationsgeräten in Betracht zieht. Wiederum ein anderes Beispiel stellt ein Diplomand dar, der seine Informationssuche in Richtung seines Diplomarbeitsthemas lenken wird. Unabhängig aus welcher Motivation heraus das Informationssystem zur Anwendung kommt, versucht der Anwender effizient an seine Informationen zu gelangen. Wie effizient in der Praxis eine Anwendung durchgeführt wird, hängt insbesondere von der zur Anwendung kommenden Technologie und von ihrer Kompatibilität oder Verträglichkeit untereinander ab. Systeme können Informationen austauschen, wenn sie den einheitlichen Standards entsprechen und untereinander kompatibel sind.

Die Interoperabilität im Sinne der Informationstechnologie stellt bedeutendes Komplement im Zusammenhang mit der Analyse und Bewertung der Einflussfaktoren dar und bietet ein Spektrum für die weitere Forschung im Bereich der informationstechnologischen Instrumente und ihrer ökonomischen Verwendbarkeit.

2.3.2 Technologie mit maschinen-verarbeitbarer Semantik

Zu Beginn des *Electronic Commerce* (B2C-Entwicklung) wurden *ShopBots* entwickelt, die verschiedene Shops besuchen, relevante Produktinformationen extrahieren und dem Web-Konsumenten einen Überblick über das vorhandene Marktangebot präsentieren. Einige Beispiele für solche im Web vorfindbaren Kaufagenten sind *mySimon.com*, *yahoo.com* und *lycos.com*, die integraler Bestandteil von Webportalen sind. Die letztgenannten bieten ihren Nutzern eine *Best-Price-Buy-Strategie* an, das heißt die Produkte werden lediglich anhand der Preise verglichen. Das einzelne *ShopBot* entnimmt die relevanten Produktpreise (entsprechend der gewünschten Kategorie) aus mehreren zur Verfügung stehenden Händlerdatenbanken [Klus01]. Die Funktionalität des *ShopBots* basiert auf *Wrappers*, die für jeden Online Shop erneut geschrieben werden. Solche *Wrappers* verwenden Keyword-Suche um auf diese Weise Informationen über das angeforderte Produkt zu finden und dem Konsumenten auf seine vorangegangene Anfrage hin zu liefern. Laut Fensel et al. [FHLW03] stößt diese Technologie dabei an folgende ernsthafte Grenzen:

- *Zum einen die Leistung:* Für jeden einzelnen Online Shop einen *Wrapper* zu schreiben erfordert viel Zeit. Auch die häufige Aktualisierung der *Wrappers* trägt zum erhöhten Aufwand bei.
- *Zum anderen die Qualität:* Die von *ShopBots* extrahierbare Produktinformation sind fehleranfällig, unvollständig und beschränkt, meistens handelt es sich dabei um Preis-Informationen. Zum Beispiel kann der *Wrapper* den Produkt-Preis extrahieren aber indirekte Kosten, wie Versandkosten werden vernachlässigt.

Diese Probleme entstehen dadurch, dass die meisten Produktinformationen auf Webseiten in natürlicher Sprache zur Verfügung stehen und weil die automatische Text-Erkennung noch immer ein Forschungsgebiet mit ungelösten Problemen ist. Hierbei ist eine *maschinen-verarbeitbare* bzw. *maschinen-interpretierbare Semantik* für Informationen erforderlich. Laut Fensel et al. [FHLW03] wird sich die Situation, sobald standardisierte Repräsentationsformen für Struktur und für die Semantik der Daten zur Verfügung stehen, drastisch ändern. Es werden *Software Agents* gebildet, die die Produktinformation der Webseiten *verstehen* können. Meta-online Shops könnten mit wenig Aufwand konstruiert werden. Diese Technik wird auch vollständige Markttransparenz in verschiedenen Dimensionen ermöglichen [FHLW03].

3 EINFLUSSFAKTOREN

Die Informationstechnologie sorgt für ein informationsgefülltes Web. Durch die Entwicklung und Realisierung neuer Technologien im Web lässt sich dieses informationsgefüllte Web besser strukturieren. Die Forschung und Entwicklung in der Technologie des Webs bieten neue Perspektiven und neue Herausforderungen an. Der Einsatz informationstechnologischer Instrumente ermöglicht zudem die Realisierung von *personalisiertem Web*. Der Anspruch an das aktuelle Web liegt in der Optimierung der Technologie und der Qualität der Informationen, und in der Umsetzung der Personalisierung im Web. Somit ergeben sich drei Einflussfaktoren des Webs, welche erst mit dem Einsatz von entsprechenden informationstechnologischen Instrumenten, optimiert werden können: „Technologie für das Web“ durch die Realisierung von Semantic Web, „Information aus dem Web“ durch den Einsatz von Semantic Web Mining und „Personalisierung im Web“ durch den Einsatz von Personalisierungssoftware.

Im Subkapitel 3.1 wird die *Technologie des Semantic Web* zum Gegenstand. Dabei wird das semantische Web mit den notwendigen Bestandteilen laut der Architektur nach Tim Berners-Lee beschrieben. Für die Realisierung des semantischen Webs sind Ontologiesprachen zu verwenden, welche daher auch zum wesentlichen Bereich des Kapitels zählen. Im Subkapitel 3.2 werden die Methoden zur Filterung von *Informationen aus dem Web* erläutert. Aus der historischen Perspektive baut das Web Mining auf Data Mining auf, so dass im Zusammenhang mit dem Semantic Web Mining beide Ansätze des Mining beschrieben werden. Das Subkapitel 3.3 erläutert anhand von Customer Relationship Management und der Recommender Systeme die Realisierung von *Personalisierung im Web*. Die folgenden Subkapitel geben somit einen Einblick in die Bestandteile und Funktionsweisen der genannten Einflussfaktoren und ihrer informationstechnologischen Instrumente.

3.1 Technologie für das Web

Das Web ist inzwischen zum alltäglichen Marktmedium geworden. Die Unternehmen kreieren eigene Webseiten um sich im neuen Marktmedium zu präsentieren und zu positionieren. Die Webseite des Unternehmens präsentiert das Unternehmensprofil, das Geschäftsfeld und in manchen Fällen auch den bisherigen Erfolg sowie Präferenzen der bisherigen namhaften Geschäftspartner. Andere Webseiten sind wiederum ausschließlich dem Zweck des Absatzes gegründet. Ein Webshop sowie alle notwendigen Informationen zur Präsentation des gesamten Angebots stellen die Mindestanforderungen an solche Webseiten dar. Der Einsatz von Webshops ist auf Webseiten, wie *amazon.com*, *al.net* oder *t-mobile.at* erfolgreich realisiert. Auch die Universitäten und die Wissenschaft nutzen das Potential dieses Marktmediums. Studierende und Interessenten aus der ganzen Welt können sich die notwendigen Informationen relativ schnell verschaffen und auf aktuellem Stand halten. Zwecks Tourismus sehen auch die meisten Hauptstädte und zahlreiche Bundesstädte die erforderliche Notwendigkeit der Präsentation im Marktmedium Web. Haasis et al. [HaSZ02] verwenden für das vorliegende wirtschaftliche Phänomen den Begriff der *digitalen Wertschöpfung*.

Tabelle 1 stellt anhand von zwei Ebenen einen Überblick über die angebotenen Dienste im Web dar. Die Ebene der öffentlichen Dienste und die Ebene der privaten Dienste sind nach den Anbietern der entsprechenden Dienste benannt. Die genannten Dienste erfüllen bedeutende Aufgaben im Tourismus, Informations-, Wirtschafts- und Bildungsbereich.

Ebenen	Anbieter	Zweck
Öffentliche Dienste	Öffentliche Behörden (Gemeinde, Bund, Länder)	Tourismus, Information, Bildung
Private Dienste	Unternehmen, Universitäten, Privatpersonen	Wirtschaft, Tourismus, Information, Bildung

Tabelle 1 Anbieter der Web-Dienste

In diesem Kapitel wird auf die wirtschaftliche Bedeutung des Markmediums Web eingegangen, da das Verständnis dessen die Grundlage für den praxisorientierten Teil dieser Arbeit darstellt. Die wirtschaftliche Bedeutung des Webs und damit auch die Bedeutung des Einflussfaktors „Technologie für das Web“ werden anhand des Ansatzes der *digitalen Wertschöpfung* nach Haasis et al. [HaSZ02] in Kapitel 4 analysiert und auf Basis dessen erweitert.

Digitale Wertschöpfung

Unter digitaler Wertschöpfung sind definitionsgemäß wirtschaftliche Tätigkeiten zu verstehen, die entweder für oder mittels elektronischer Netze und Medien betrieben werden [HaSZ02]. Der branchenübergreifende Einsatz digitaler Medien für die Geschäftstätigkeit, wird auch als *Electronic Business* bezeichnet [HaSZ02].

Die digitale Wertschöpfung tritt in *verschiedenen Lebensbereichen* als wirtschaftliches Phänomen auf. Somit lässt sich die Gliederung der zwei genannten Ebenen erweitern: Beispiele für mögliche Lebensbereiche sind Bildung (*Telelearning, virtuelles Studium*), Kultur (*Medienkunst*), Verwaltung (*digitale Behörde*) und Wirtschaft (*Unternehmen, öffentliche Behörden, Private*) [HaSZ02]. Diese Bereiche werden laut Haasis et al. [HaSZ02] für *Dimensionen der digitalen Wertschöpfung* erklärt.

Die digitale Wertschöpfung ist insbesondere in der Zeit der Globalisierung und des international wachsenden Wettbewerbs ein entscheidender Faktor [HaSZ02]. Das Potential der digitalen Wertschöpfung lässt sich durch verbesserte Webtechnologie genauer realisieren, da die wirtschaftlichen Tätigkeiten durch Verbesserung der Webtechnologie ebenfalls effizienter verrichtet werden können. Mit der zunehmenden Bedeutung als wirtschaftlicher Faktor gewinnt auch die Forschung und Entwicklung dieser Technologie an Bedeutung. Im Zusammenhang mit der Informationsgewinnung und gemeinsamen Nutzung von Informationen entsteht der Anspruch auf Kompatibilität der Daten. Die richtige Verwendung von Daten und Informationen setzt voraus, dass den Daten auch die richtige Bedeutung zugewiesen wird und in weiterer Folge die Informationen nicht nur vom Menschen sondern auch von der Maschine richtig interpretiert werden können.

Um Daten eine Bedeutung zu verleihen, muss das Wissen auf eine Weise repräsentiert werden, die von der entsprechenden Sprache abhängt. In diesem Zusammenhang ist der Bereich *Artificial Intelligence* zu nennen, der sich mit Repräsentation von Wissen befasst. Das Ziel besteht darin, Strukturen zu schaffen, die es erlauben Informationen effizient zu lagern, effizient zu modifizieren und logisch aufzubauen [FHLW03]. Beispiele solcher Sprachen, die das Wissen repräsentieren, sind KRL [BoWi77], KL-ONE [BrSc85], Cyc-L [LeGu90], LOOM [MaBr92] und Classic [McRI95].

Einer der ältesten Formalismen für die Repräsentation von Wissen stellt das *Semantic Networks* dar [FHLW03]. Ein *Semantic Network* repräsentiert Wissen in Form von Knoten, die mit entsprechenden Kanten in Abhängigkeit ihrer Zugehörigkeit verknüpft sind. In dieser Repräsentation wird die Bedeutung impliziert, dass ein Begriff mit anderen Begriffen in Beziehung gesetzt wird, indem es Eigenschaften erbt bzw. vererbt. Die Knoten sind die Begriffe und Kanten die Beziehungen zwischen den Knoten. Das themenspezifische und zugleich abstrakte Beispiel kann das *World Wide Web* und *Semantic Web* genannt werden, welche als Knoten definiert und durch Kanten in Beziehung gesetzt werden können. Das bereits bestehende *World Wide Web* ist durch bestimmte Merkmale charakterisiert, welche an *Semantic Web* weitervererbt werden. Diese Merkmale sind zum Beispiel der weltweite Informationsaustausch über das Netz oder die Intensivierung der Kommunikation zwischen den Menschen. Als eine Art der Modifikation im Sinne einer Erweiterung weist das *Semantic Web* zu den bereits geerbten Merkmalen zusätzliche Charakteristika auf. Als Beispiel für besonderes Merkmal des *Semantic Web* kann die verbesserte Interaktion zwischen Mensch und Maschine genannt werden, da die Maschine im semantischen Web im Stande ist Informationen genauer zu interpretieren. Während die Informationen im aktuellen Netz lediglich für Menschen verständlich sind, sind diese Informationen im semantischen Netz in einer Sprache des semantischen Webs entsprechend definiert und somit auch für die Maschine interpretierbar. Die Sprachen des *Semantic Web* werden in der Literatur auch Ontologiesprachen genannt, da sie die Begriffe definieren bzw. ihnen eine Bedeutung zuordnen.

Frame Systems ist eine andere Repräsentationsart und im Verhältnis zu *Semantic Networks* von gleicher Form (isomorph). In der Terminologie der *Frame Systeme* stellt der Rahmen ein Objekt von Daten dar, das aus Vertiefungen besteht. Dabei repräsentiert jede Vertiefung eine Eigenschaft oder ein Attribut des Objekts. Vertiefungen können einen Wert oder mehrere Werte haben. Diese Werte können wiederum Verweise auf andere *Frames* sein [FHLW03].

Die Bedeutung des Marktmediums Web wird anhand der vorgestellten digitalen Wertschöpfung begründet. Der Großteil an wirtschaftlichen Tätigkeiten kann erst durch das Web realisiert werden, nämlich immer dann wenn der potentielle Kunde nur via Web kontaktiert werden kann oder wenn das Unternehmen sein Service und Angebot ausschließlich via Web betreibt. Letzteres kann auf Grund wirtschaftlicher Vorteile profitabler sein, als die Eröffnung von unzähligen Filialen. Ein anschauliches Praxisbeispiel stellt der Vertrieb von mobilen Telekommunikationsgeräten mithilfe von Online-Shops dar. Zur Servicierung von Kundenanfragen wird aufgrund von Outsourcing-Vorteilen in der Praxis ein Call- & Communication Center beauftragt. Ein Beispiel für ein solches Unternehmen ist das *Walter MasterManagement GmbH*.

Der Aufbau einer Webseite trägt zum Zufriedenheitsgrad des Anwenders bei. Ist die Frontpage überschaubar und klar strukturiert, wird auch der Anwender rascher zum gesuchten Produkt gelangen. Ein Überblick über eine internationale Webseite wird anhand von folgenden Dimensionen dargestellt: *Sprachen, Themen, Anwenderebenen, kundenspezifischer Bereich, kundenspezifische Leistungen*. Einige Beispiele für diese Dimensionen werden im Folgenden aufgelistet:

1. Sprachen (Englisch, Deutsch, Spanisch, Slowenisch, Tschechisch, Kroatisch, etc.)
2. Themen (Unternehmensprofil, Service & Angebot, Unternehmenserfolg, etc.)
3. Anwenderebenen (Privat, Business)
4. Kundenspezifischer Bereich (Login, Passwort)
5. Kundenspezifische Leistungen (branchenspezifisch, unternehmensspezifisch)

Welche Leistungen dem Kunden angeboten werden bzw. welcher Kunde in den Genuss bestimmter Leistungen kommt, ist sowohl branchenspezifisch (das heißt von der Branche des entsprechenden Unternehmens abhängig) als auch unternehmensspezifisch (das heißt von der Unternehmenspolitik abhängig). Ein Unternehmen, das gemäß *Customer Relationship Management* die Beziehung zu seinen Kunden pflegt, wird insbesondere spezielle Angebote und Prämien für seine treuen Kunden bereitstellen.

Im folgenden Subkapitel wird das Konzept des *Semantic Web* vorgestellt. Das semantische Web ist die Idee des World Wide Web-Erfinders *Tim Berners-Lee*. Dieses Konzept stellt eine Erweiterung des *World Wide Web* um maschinenlesbare Daten dar. Die Semantik wird in diesem Konzept formal anhand maschinenlesbarer Daten festgelegt. Nach diesem Konzept stehen Informationen nicht nur in der für Menschen lesbaren Form, sondern insbesondere auch formal, in einer für maschinen-interpretierbaren Form zur Verfügung [Wiki05].

3.1.1 Semantic Web

„The Semantic Web is an extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation.“ [BeHL01]

Die Definition von Tim Berners-Lee et al. [BeHL01] hebt deutlich das primäre Ziel des *Semantic Web* hervor: bessere Zusammenarbeit zwischen Mensch und Maschine. Das Ziel wird anhand maschinen-verarbeitbarer Semantik verwirklicht. Diesem Zweck werden Sprachen entwickelt und in der Praxis eingesetzt, die bestimmte Voraussetzungen erfüllen. Im Folgenden wird auf die Eigenschaften von *Semantic Web Sprachen* eingegangen.

Sprachen für *Semantic Web* haben folgende Aspekte zu erfüllen: sie müssen über *formale Syntax* und *formale Semantik* verfügen um automatisierte Verarbeitung ihrer Inhalte zu gewährleisten bzw. erst zu ermöglichen. Sie müssen ebenso über *standardisiertes Vokabular*, das auf *Real-world Semantik* bezogen ist, verfügen. So wird der effiziente Austausch von Informationen und Wissen zwischen Mensch und Maschine ermöglicht [FHLW03].

Der Gründer des World Wide Web, Tim Berners-Lee, entwickelte die grundlegende Idee für die Weiterentwicklung des Webs. Dazu erstellte er die siebenstufige Architektur des Semantic Web, welche keinesfalls eine Gebrauchsanweisung für die Entwicklung des semantischen Webs darstellt sondern lediglich als Vorschlag dient. Tabelle 2 zeigt die von Tim Berners-Lee erstellte Architektur des semantischen Webs. Basierend auf *Unicode* (<http://www.unicode.org>) und *Unified Resource Identifier* (URI) werden Sprachen wie *eXtensible Markup Language* (XML) und *Resource Description Framework* (RDF) eingesetzt. Nachfolgend wird die Architektur mit ihren Bestandteilen erläutert.

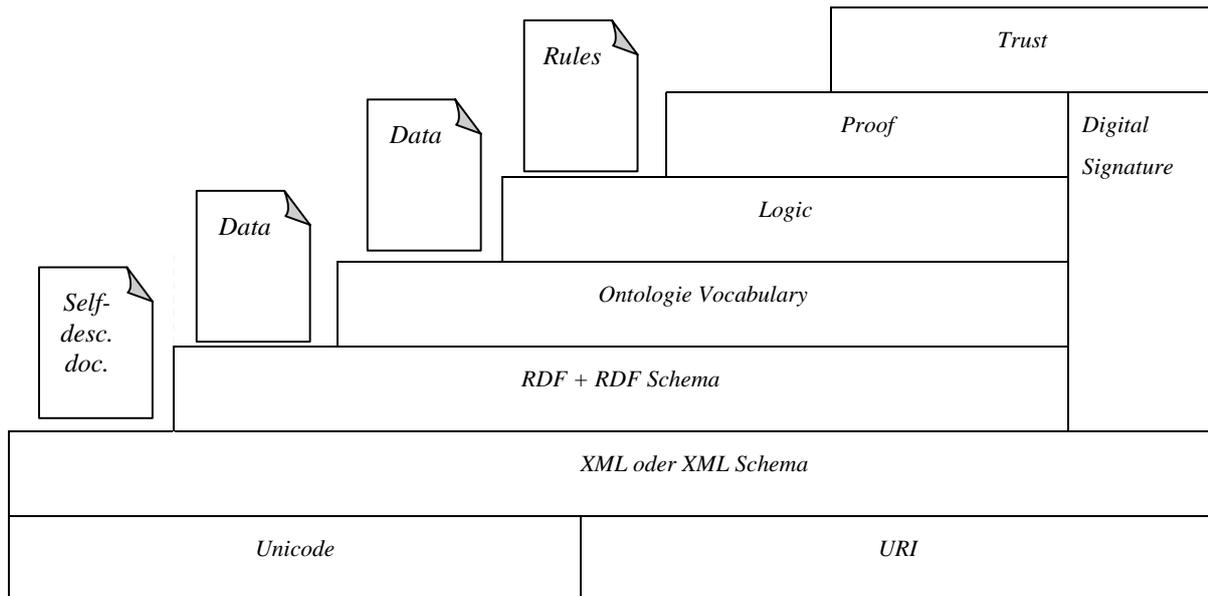


Tabelle 2 Berners-Lee's Architektur des Semantic Web [McGu02]

Die Stufen der Berners-Lee's Architektur sind so angeordnet, dass die jeweils obere Stufe eine Teilmenge von der Unteren ist und daher die Reihenfolge eingehalten werden muss, wenn eine stabile Ontologiumgebung aufgebaut werden soll. Die oberen Stufen machen von Unterliegenden Gebrauch [Eber04]. Zunächst wird mit *Unicode* bzw. *URI* [erste Stufe] abstrakt betrachtet ein Raum im Web begründet, worauf mit Auszeichnungssprachen ganze Webseiten mit spezifischen Inhalten entstehen können. Durch den Einsatz bestimmter Web-Sprachen (auch Ontologiesprachen genannt) wird es möglich Daten nicht nur zu repräsentieren sondern insbesondere für Maschinen verarbeitbar bzw. interpretierbar zu gestalten. Dabei wird die Syntax der Auszeichnungssprache, *XML* bzw. *XML Schema* [zweite Stufe], für den effektiven Einsatz von *RDF* bzw. *RDF Schema* [dritte Stufe] verwendet. Daher baut die RDF-Stufe auf die XML-Stufe auf und nicht umgekehrt. Die Dokumentation der eigenen Darstellungsform der Daten [*Self-description documentation* bzw. *self-desc. doc.*] wird von W3C Consortium empfohlen. Anderen Literaturquellen [vgl. Eber04] zufolge ist die Dokumentation in diesem Zusammenhang optional, wenn die Darstellungsform die der RDF entspricht. In einem Netz mit maschinen-verarbeitbaren Informationen ist ein entsprechendes Vokabular erforderlich, das *Ontologie Vokabular* [vierte Stufe], welches mit einem Katalog oder Glossar vergleichbar ist. Das Ontologie-Vokabular enthält Begriffe mit der dazugehörigen Bedeutung. Ontologie selbst ist in der Literatur des semantischen Webs ein weit verbreiteter Begriff und wird daher in Kapitel 3.1.2 im Detail behandelt.

In der Berners-Lee's Architektur sind auf den einzelnen Stufen zusätzliche Blätter angeordnet. Da für das Ontologie-Vokabular bestimmte Begriffe erforderlich sind, die sich aus verschiedenen Datensätzen zusammensetzen, werden die *Daten* [*Data*] als eigene Blätter in der Architektur dargestellt. Auf die Stufe

des Ontologie-Vokabulars bauen in der Architektur die Stufen der *Logik [fünfte Stufe]*, des *Beweises [sechste Stufe]* und die *Stufe des Vertrauens in das System [siebente Stufe]* auf.

Aus einfachen Daten [*Data*] wird mithilfe von Ontologie-Sprachen (zum Beispiel RDF) ein maschineninterpretierbares Ontologie-Vokabular erzeugt. Diese maschinen-verarbeitbaren Informationen können nach Grundsätzen der *Logik [fünfte Stufe]* von Suchmaschinen aussortiert werden, wenn sie zudem auch dem *Beweis [Proof, sechste Stufe]* standhalten. Das heißt, wenn die Informationen ihre Gültigkeit auch nach der Überprüfung bestimmter Regeln [*Rules*], zum Beispiel Deduktionsregeln, behalten. Parallel dazu kann die Technologie der Public Key Kryptographie dazu verwendet werden um bestimmte Texte in Geheimentexte zu verschlüsseln, welche wiederum nur mit dem geheimen Schlüssel wieder entschlüsselt werden können. Somit baut das Prinzip der *elektronischen Signatur [Digital Signature]* parallel mit den bereits genannten Stufen *Vertrauen in das System [siebente Stufe]* seitens der Anwender auf.

Zur Verwirklichung der fünften Stufe [*Logic*] bedarf es entsprechender Ontologiesprachen, welche anschließend als Sprachen der 2. Web-Generation beschrieben werden. Die Stufe des Beweises und des Vertrauens sind langfristige Forschungsziele, welche sich erst mit den zukünftigen Forschungsergebnissen optimal verwirklichen lassen [Eber04].

Nachfolgend wird näher auf die unterste Schicht der Berners-Lee's Architektur eingegangen. Unicode und URI werden daher näher erläutert.

Da Computer grundsätzlich mit Zahlen arbeiten, werden Buchstaben und andere Zeichen den Zahlen zugeordnet. Inzwischen wurden unterschiedliche Kodierungssysteme zwecks dieser Zuordnung entwickelt, wie zum Beispiel *American Standard Code for Information Interchange* (ASCII) oder *Code of American National Standards Institute* (ANSI-Code). In der Praxis benötigt allein die Europäische Union mehrere Kodierungssysteme um damit den Bedarf für die Sprachen aller Mitgliedsländer abzudecken. Unterschiedliche Kodierungssysteme können untereinander dieselbe Zahl für verschiedene Zeichen benutzen oder verschiedene Zahlen für dasselbe Zeichen. Daran lässt sich die Unverträglichkeit der Kodierungssysteme untereinander erkennen. Das Kodierungssystem *Unicode* gibt erstmals jedem einzelnen Buchstaben, Interpunktionszeichen und jedem einzelnen technischen Zeichen eine eigene Nummer. Zudem ist *Unicode* plattformunabhängig, programmunabhängig und sprachenunabhängig [Unic06].

Mit *Uniform Resource Identifier* (URI) wird in Tabelle 2 die Adresse einer Webseite bezeichnet. Ein Beispiel für URI ist die Webseite der Wirtschaftsuniversität Wien, <http://www.wu-wien.ac.at/>. URI ist zugleich eine Identifizierungsadresse wodurch es sich von *Unified Resource Locator* (URL) unterscheidet. So kann die Webadresse der Wirtschaftsuniversität Wien als Identifizierungsadresse für andere Webseiten wie zum Beispiel jene der Studierenden oder jene der Assistenten und Professoren an der Wirtschaftsuniversität Wien darstellen. Ein anderes Beispiel für eine URL ist die Webseite vom Zentrum für Auslandsstudien, <http://www.wu-wien.ac.at/zas>.

Die genannten Kodierungssysteme und die Adressen der Webseiten bilden somit den Grundstein für das *Semantic Web*. Damit nun vom *Semantic Web* auch tatsächlich gesprochen werden kann, werden die für das semantische Web erforderlichen Sprachen bzw. weitere Bestandteile aus der Berners-Lee's Architektur näher erläutert (wie XML, RDF und Ontologien). Vorweg wird auf die Sprachen der ersten Web-Generation eingegangen, da diese die Vorreiter der zweiten Web-Generation darstellen.

Sprachen der 1. Web-Generation

Fensel et al. [FHLW03] liefern im Verhältnis zur bereits vorgestellten Berners-Lee's Architektur etwas detailliertere Darstellungsform für Websprachen. Während die Architektur von Berners-Lee ausschließlich jene des semantischen Webs darstellt, ist die Sprachenarchitektur nach Fensel et al. [FHLW03] umfassender, indem sie auch die Sprachen des noch vorherrschenden *World Wide Web* (zum Beispiel HTML, XML) inkludiert. Bild 1 zeigt die alternative Darstellung der Sprachenarchitektur nach Fensel et al. [FHLW03]. Die tatsächliche Sprachenarchitektur des semantischen Webs ist somit erst mit der übergeordneten Schicht, das heißt mit RDF im Bild 1, gegeben. Das Besondere an der Sprachenarchitektur nach Fensel et al. [FHLW03] ist, dass es sowohl *Sprachen der ersten Web-Generation* als auch *Sprachen der zweiten Web-Generation* präsentiert und alle diese Sprachen durch das gemeinsame Merkmal der Auszeichnungssprachen (*Markup Languages*) charakterisiert sind. Die einzelnen Schichten dieser Architektur werden anschließend beschrieben.

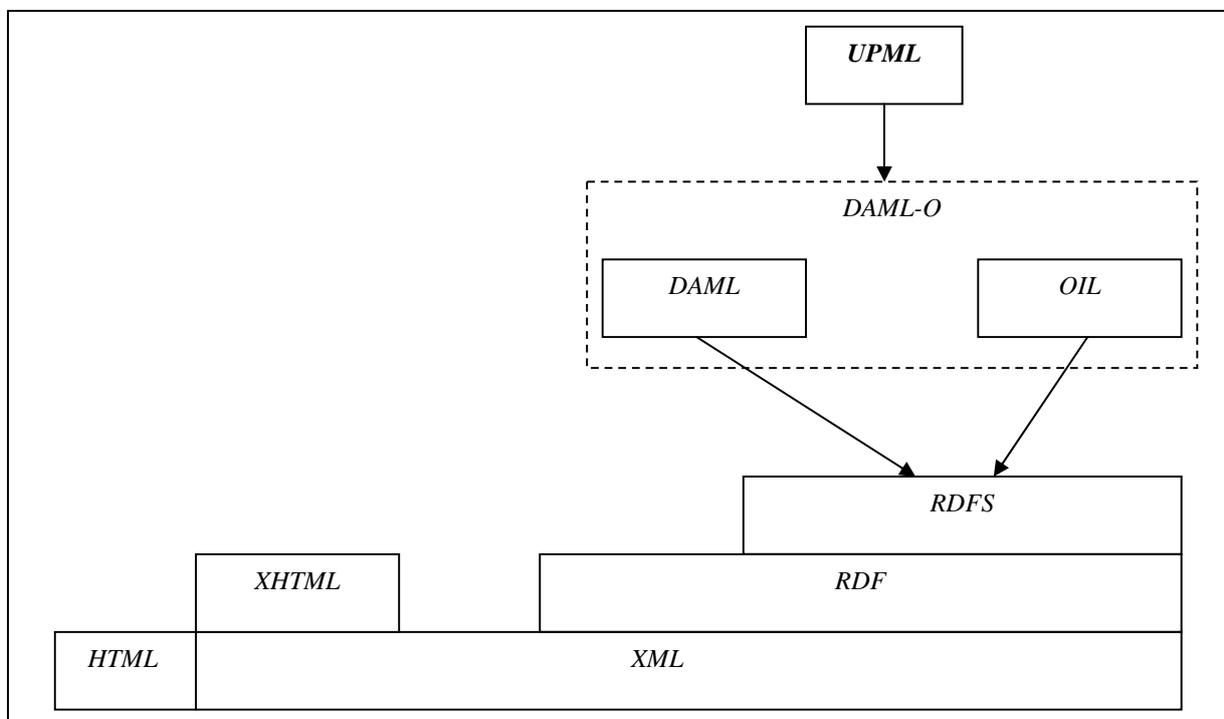


Bild 1 Sprachenarchitektur nach Fensel et al. [FHLW03]

Die im Bild 1 als unterste dargestellte Schicht, wird von *HyperText Markup Language* (HTML) und *eXtensible Markup Language* (XML) gebildet. HTML ist eine Auszeichnungssprache zur Beschreibung von Informationen in Hypertext (bzw. Hyperlinks), welche Anfangs der 90er Jahren des 20. Jahrhunderts das *World Wide Web* erst ermöglichten. Die Komplexität des *Standard Generalized Markup Language* (SGML) erschwerte die Weiterentwicklung der Software, welche daher auch der unpopuläre Vorgänger von HTML blieb. Erst die Einfachheit von HTML ermöglichte das rasante Wachstum des *World Wide Webs*. Mit HTML können Informationen im Web einfach repräsentiert und formatiert werden. Jedoch hat in weiterer Folge gerade diese Einfachheit den weiteren Fortschritt der Web-Applikation in vielen Domains und bei vielen Aufgaben behindert, da HTML aufgrund seiner Unflexibilität auf die

Repräsentation und Formatierung der Daten beschränkt war. Dies war der ausschlaggebende Grund eine neue Sprache zu definieren, eine verbesserte Sprache, XML [FHLW03].

Die *eXtensible Markup Language* oder auf Deutsch erweiterbare Auszeichnungssprache dient standardmäßig der Erstellung maschinen- und menschenlesbarer Dokumente in Form einer Baumstruktur. Die Regeln für den Aufbau solcher Dokumente werden von XML vordefiniert, wobei für jeden konkreten Einzelfall die Details des Dokuments gesondert spezifiziert werden müssen, insbesondere die Festlegung und Anordnung der Strukturelemente (XML-Elemente) innerhalb des Dokumentenbaums. Die Namen der Strukturelemente können in XML frei gewählt werden [Wiki05].

Die Sprachen der zweiten Schicht, das heißt ab *Resource Description Framework* (RDF) gehören bereits zu den Sprachen des *Semantic Web*, welche daher unter den Sprachen der zweiten Web-Generation präsentiert werden.

Sprachen der 2. Web-Generation

Die entwickelten Sprachen für das semantische Web werden auch Ontologiesprachen genannt, da sie insbesondere die erforderliche Definition der Begriffe und ihre Semantik voraussetzen. Jede Ontologiesprache muss laut Fensel et al. [FHLW03] drei bedeutende Anforderungen erfüllen:

- Sie muss hoch intuitiv zum aktiven Benutzer sein.
- Sie muss wohldefinierte, formale Semantik aufweisen
- Sie muss einen geeigneten Link mit Web-Sprachen wie XML und RDF beinhalten um die Interoperabilität zu gewährleisten.

Die Weiterentwicklung der vorhandenen Sprachen durch das *World Wide Web Consortium* (W3C, <http://www.w3.org>) führte zu neuen Sprachen, wie *eXtensible HyperText Markup Language* (XHTML) und *Resource Description Framework* (RDF). XHTML ist eine HTML ähnliche Sprache, die den Regeln der XML-Syntax entspricht. Grundlegender Unterschied zwischen HTML und XHTML liegt darin, dass HTML direkter Nachfolger von SGML ist, während XHTML auf Basis von XML definiert wurde und somit den direkten Vorgänger XML hat. Dieser Sachverhalt wird zwecks Übersicht im Bild 2 veranschaulicht. XML wiederum ist ebenfalls direkter Nachfolger von SGML sowie direkter Vorgänger von XHTML und ein potentieller Vorgänger von RDF.

Resource Description Framework (RDF) wird auch als das Model zur Repräsentation von Metadaten genannt [W3C99]. Hierbei ist die Unterscheidung zwischen den Daten, die maschinen-lesbar (*machine-readable*) und die maschinen-verständlich (*machine-understandable*) sind, von Bedeutung. Die meisten Daten, die zur reinen Darstellung dienen, werden von Maschinen *nicht verstanden*. Es ist schwierig im Web alles zu automatisieren. Aufgrund des Informationsvolumens im Web ist auch die manuelle Automatisierung zeitintensiv. Daher wird in diesem Zusammenhang der *Lösungsansatz der Metadaten* präsentiert. Metadaten sind sozusagen *Daten über Daten*. So ist ein Katalog der Bibliothek eine mögliche Form der Metadaten, da es Publikationen beschreibt. Im Kontext von Spezifikationen sind Metadaten jene Daten, die die Web-Ressourcen beschreiben [W3C99].

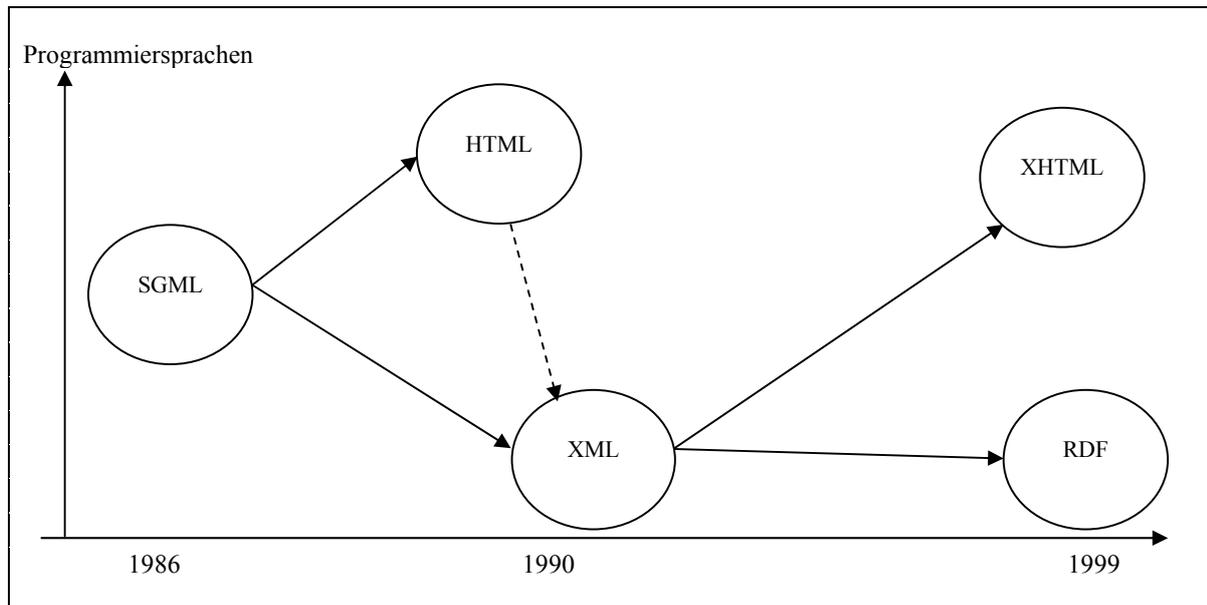


Bild 2 Entwicklung der Web-Sprachen

Resource Description Framework (RDF) dient der Aufbereitung und Entwicklung von Metadaten. RDF gewährleistet Interoperabilität zwischen Anwendungen, die maschinen-verständliche Informationen im Web austauschen. Insbesondere wird durch RDF automatische Aufbereitung von Web-Ressourcen realisiert. In Vielzahl von Anwendungsgebieten kann RDF eingesetzt werden, wie zum Beispiel für die Beschaffung von Informationen durch verbesserte Leistung der Suchmaschinen, für Kataloge um den Inhalt und die damit verbundenen Beziehungen auf einer bestimmten Webseite zu beschreiben oder für Digitale Bibliotheken durch intelligente Software Agents um Informationsaustausch zu ermöglichen. Auch für Bewertungsverfahren lässt sich RDF einsetzen oder für die Beschreibung der intellektuellen Eigentümerrechte von Webseiten bzw. zur Darstellung von Datenschutzaspekten [W3C99].

Die Syntax des RDF nach der W3C-Empfehlung verwendet XML. Das Ziel des RDF ist es die Semantik der Daten, die auf XML basieren, auf standardisierte und vollständig kompatible Weise zu spezifizieren. RDF und XML sind soweit komplementär zueinander, dass sie sich gegenseitig ergänzen können, wobei die XML-Syntax nur eine mögliche Variante für RDF darstellt und die Entwicklung anderer Darstellungsformen für RDF Daten Modelle nicht ausgeschlossen sind [W3C99].

RDF hat ein Klassensystem wie die meisten objektorientierten Programmiersprachen und Modellierungssysteme. Eine Sammlung an Klassen nennt man Schema, daher spricht man in weiterer Folge auch von *Resource Description Framework Schema* (RDFS). Klassen sind in einer Hierarchie organisiert und bieten Erweiterungsmöglichkeiten durch die Verfeinerung in Subklassen an. Um anhand eines existierenden Schemas ein ähnliches Schema zu erzeugen, muss das Rad nicht neu erfunden werden. Es reicht eine schrittweise zunehmende Modifizierung (*inkrementelle Modifizierung*) anhand von Basis-Schema. Durch die simultane Benutzbarkeit (*Mehrbenutzbarkeit*) von Schemas, kann RDF die Wiederverwendbarkeit von Definitionen der Metadaten unterstützen. Die Wiederverwendbarkeit und inkrementelle Erweiterbarkeit des RDF erlauben die Verwendung von mehrfacher Vererbung. RDF kann als einfaches Frame System charakterisiert werden, auf welches der Beweismechanismus aufgebaut werden kann [W3C99]. RDFS bildet die dritte Schicht der Sprachenarchitektur nach Fensel et al.

Ontology Inference Layer bzw. *Ontology Interchange Layer* (OIL) ist eine Sprache, welche die webbasierte Repräsentation von Ontologien darstellt und Framebased Systeme, XML und RDF, verwendet. OIL findet Anwendung insbesondere in Einsatzbereichen der Suchmaschinen, Wissensmanagement und E-Commerce.

Darpa Agent Markup Language (DAML) wurde von der *Defence Advanced Research Projects Agency* (DARPA), der zentralen Forschungs- und Entwicklungsorganisation des Verteidigungsministeriums der USA, entwickelt [FHLW03]. Wird hierbei eine Repräsentation gewählt, die auf dem *Ontology Inference Layer* basiert, spricht man von *Darpa Agent Markup Language for Ontologies* (DAML-O). DAML-O ist eine Ontologiesprache, welche maschinen-verständliche Semantik gewährleistet um somit auch die Kommunikation zwischen Mensch und Maschine zu verbessern. DAML-O und OIL bilden die vierte Schicht der Sprachenarchitektur nach Fensel et al.

Unified Problem-solving Method-development Language (UPML) entstand aus dem IBROW-Projekt der Stanford University (Project on *Intelligent Brokering Service for Knowledge-Component Reuse on the World Wide Web*). UPML basiert auf einer kompetenzmodellierten Architektur und unterstützt die Argumentation der intelligenten Vermittler (*intelligent brokers*), welche wiederverwendbare Softwarekomponenten dezentralisiert konfigurieren können. Dabei legt der Benutzer die Ziele seiner Aufgaben fest, welche vom Vermittler übernommen werden. Der Vermittler greift nun auf dezentralisierte digitale Archive von wieder verwendbaren Problemlösungsmethoden im Web zu, wählt jene Komponenten aus welche der jeweiligen Anfrage entsprechen. Nachdem die Komponenten den Anfragen angepasst wurden, werden sie konfiguriert. Zudem stellt der Vermittler eine Beschreibung der durchführbaren Komponentenkonfiguration bereit. Praxisbeispiele für Archive solcher Softwarekomponenten sind Klassifikation, Dokumentenanalyse, Informationsfilter, fallgestützte Argumentation und Planung. Zudem wurde eine Vielzahl von Nebenanwendungen, wie zum Beispiel ein Webinformationsvermittler, ein Gesundheitsfürsorgedienst, eine Anwendung zur Analyse und Klassifikation von Konferenzpapieren und eine Reihe von Klassifikationsanwendungen entwickelt [Wiel04]. UPML bildet die oberste Schicht der Sprachenarchitektur nach Fensel et al. Somit nimmt mit jeder zusätzlichen Schicht (von der untersten zu der obersten Schicht) auch die Spezifikation zu.

3.1.2 Ontologien

“Eine Ontologie ist eine formale, explizite Spezifikation einer gemeinsamen Konzeptualisierung.” [Grub93]

Konzeptualisierung bezieht sich auf ein abstraktes Model einiger Phänomene in der Welt, das relevante Konzepte dieser Phänomene identifiziert. *Gemeinsam* reflektiert in Zusammenhang mit der Konzeptualisierung die Auffassung, dass eine Ontologie *konsensuelles Wissen* aufbringt, das heißt es ist nicht auf ein Individuum begrenzt sondern wird von einer Gruppe vertreten. Mit *explizit* wird ausgedrückt, dass die Art der verwendeten Konzepte und der Bedingungen ihrer Verwendung ausdrücklich definiert sind. *Formal* bezieht sich auf die Tatsache, dass die Ontologie maschinen-verständlich sein sollte. Verschiedene Grade der Formalität sind durchaus vorstellbar [FHLW03].

Im Bereich der Wissenspräsentation wird unter einer Ontologie ein formal definiertes System von Begriffen, Konzepten und Relationen zwischen diesen Begriffen verstanden. Ontologien besitzen – zumindest implizit – auch Regeln [Wiki05].

Nach der Art der Repräsentation von Ontologien werden in der Literatur [Wiki05] folgende Typen unterschieden:

1. *Taxonomie* (=Systematik): Objekte werden streng hierarchisch klassifiziert, Taxonomien werden beispielsweise durch Bäume visualisiert.
2. *Thesaurus*: Objekte werden miteinander beliebig in Beziehung gesetzt.
3. *Logisch-mathematische Repräsentation*: Objektbeziehungen werden über formale Notationen dargestellt (zum Beispiel *Synonym* $(a,b) : = \text{Synonym}(b,a)$;).

Im sprachlichen Gebrauch kann eine Ontologie entweder Allgemeinwissen oder einen fachspezifischen Wissensbereich abbilden. Die Informatik liefert Ansätze zur automatischen Unterstützung der Ontologieerstellung [FaSt02].

Laut der Definition von Protégé-Team stellt eine Ontologie das Vokabular für die Domäne zur Verfügung inklusive der Bedeutungen dieser Begriffe und beschreibt die Konzepte und Beziehungen, die in dieser Domäne relevant sind [Stan05].

Die bereits vorgestellten Auszeichnungssprachen für das semantische Web (RDF, DAML-O, OIL und UPML) dienen der Beschreibung von Ontologien. Aber auch *Web Ontology Language* (OWL) erfüllt diese Funktion und wird ebenfalls als Ontologiesprache definiert. *Web Ontology Language* (OWL) wurde von W3C entwickelt um Ontologien anhand einer formalen Sprache erstellen zu können. Dabei werden Begriffe einer Domäne und deren Beziehungen formal so beschrieben, dass auch Software-Agenten die Bedeutung *verarbeiten* bzw. *verstehen* können. Verstehen heißt in diesem Zusammenhang Daten und Informationen richtig zu interpretieren. OWL ist ein wesentlicher Bestandteil der Idee des semantischen Webs. OWL basiert technisch auf der RDF-Syntax und historisch auf DAML-O [Wiki05].

Eines der einfachsten Beispiele von möglichen Ontologien stellt ein kontrolliertes Vokabular dar, das ist eine begrenzte Liste von Begriffen, die sich in Katalogen widerspiegelt [FHLW03]. Kataloge verfügen über eindeutige Interpretation der Begriffe. So hat jede Verwendung eines Begriffes einen exakten Bezeichner/Identifikator, zum Beispiel 25 [FHLW03].

Eine andere Variante einer Ontologie-Spezifikation ist das Glossar, das heißt eine Liste mit Begriffen und der dazugehörigen Bedeutung. Die Bedeutung der einzelnen Begriffe ist anhand von Aussagen in natürlicher Sprache spezifiziert. Solange diese Aussagen für den Benutzer in natürlicher Sprache verständlich spezifiziert sind und somit auch interpretierbar sind, stellen sie mit ihrer Bedeutung zugleich eine Art der Semantik dar [FHLW03]. Normalerweise sind Interpretationen nicht eindeutig, so dass diese Spezifikationen für Computer-Agenten ungeeignet sind. Deshalb erfüllt die Glossar-Definition nicht die Anforderungen des maschinen-verarbeitbaren Ansatzes [FHLW03].

Thesauri verfügen über zusätzliche Semantik in ihrer Behandlung von Relationen zwischen Begriffen. Sie verfügen über Informationen wie zum Beispiel synonyme Beziehungen [FHLW03]. In vielen Fällen können die Beziehungen von Agenten uneindeutig interpretiert werden. Normalerweise verfügen Thesauri

über keine explizite Hierarchie von Begriffen, obwohl eine einfache Hierarchie mit engeren und erweiterten Begriffsspezifikationen abgeleitet werden könnte [FHLW03].

Da Ontologien mehr Informationen formulieren und daher zwecks der Erhöhung des Informationsgrades zum Einsatz kommen, steigt auch proportional das Erfordernis ihres Einsatzes [FHLW03].

Folgende notwendigen Bestandteile einer Ontologie werden laut Fensel et al. [FHLW03] genannt:

- kontrollierbares und erweiterbares Vokabular
- eindeutige Interpretation von Beziehungen zwischen Klassen und Begriffen
- strenge hierarchische Subklassen-Beziehung zwischen Klassen

Jene Spezifikationen, die über diese drei genannten Bestandteile verfügen, gehören zu der Gruppe der *einfachen Ontologien*. Zudem gibt es verschiedene Bestandteile, welche für Ontologien nicht obligatorisch sind [vgl. FHLW03].

Komplexe Ontologien sind im Vergleich zu einfachen Ontologien kostenintensiver und daher auch seltener vorzufinden. Einfache Ontologien stehen in vielen Formaten zur Verfügung: viele sind in Form von Freeware im Web downloadbar und andere existieren in Form von internen Informationsorganisationsstrukturen in Unternehmen, Universitäten und dergleichen [FHLW03].

Einfache Ontologien

Die Verwendung einfacher Ontologien findet in der Praxis aus unterschiedlicher Motivation heraus statt. Einige Möglichkeiten der Anwendung in der Praxis werden laut Fensel et al. [FHLW03] diskutiert:

- Die Ontologien stellen für die dazugehörige Domäne ein *kontrollierbares Vokabular* zur Verfügung. Dadurch können Benutzer, Autoren und Datenbanken die Begriffe von demselben Vokabular verwenden. Zusätzlich können Programme Interfaces generieren, die die Verwendung von kontrollierten Begriffen einer Ontologie fördern. Daraus lässt sich die Verwendung von gleichen Begriffsrahmen ableiten. Zwar können manche Begriffe noch immer mit unterschiedlicher Bedeutung verwendet werden, die allgemeine Begriffsverwendung begründet jedoch die Basis des Konzeptes der Interoperabilität [FHLW03].
- Eine einfache Taxonomie kann zur Organisation der Webseite und als Orientierungsbasis verwendet werden. In vielen Webseiten werden Top-Themen auf die Frontseite gestellt um die notwendige Aufmerksamkeit der Internet-Benutzer auf den Inhalt der Webseite zu lenken. So werden wesentliche Inhalte mit gemeinsamen Merkmalen zu Gruppen zusammengefasst. Diese Generalisierungshierarchie von Begriffen und Themen dient dem Überblick über den Inhalt der Webseite und stellt die notwendige Verbindung zu den Subkategorien dar [FHLW03].
- Taxonomien können auch die Erwartungseinstellungen (*Relevanzeinschätzung*) unterstützen. Wenn die Webseite für den Benutzer relevante Merkmale aufweist, kann der Benutzer anhand der

Hierarchie der Top-Themen rasch bestimmen ob die Webseite für ihn interessanten Inhalt aufweist bzw. interessante Dienstleistungen anbietet. Daher ist die sachliche Benutzeroberfläche einer Webseite wichtiges Kriterium sowie eine Voraussetzung für die richtige Einschätzung der Benutzer in Bezug auf die Relevanz des Inhalts der entsprechenden Webseite.

- Taxonomien können als eine Dachstruktur (*Umbrella Structures*) zur Anwendung kommen aus der sich der Inhalt ausbauen bzw. prolongieren lässt. *United Standard Products and Services Code* (UNSPSC, www.unspsc.org) ist eine frei zur Verfügung stehende Ontologie. UNSPSC wurde im Jahr 1998 von *United Nations Development Programme* (UNDP) und *Dunn & Bradstreet* (D & B) gegründet und sollte die Infrastruktur für Interoperabilität von Begriffen in Domänen von Produkten und Dienstleistungen darstellen [UNSP03].
- Taxonomien unterstützen auch das Browsing durch den Benutzer. Der Inhalt der Webseite kann mit den Begriffen der Taxonomie markiert bzw. identifiziert werden. Dies kann entweder manuell im Yahoo-Stil oder automatisch anhand der Verwendung vom Clustering-Ansatz durchgeführt werden. Wird eine Webseite mit den Begriffen, die aus dem kontrollierten Vokabular, meta-markiert, dann können Suchmaschinen die Identifikation auswerten. Somit unterstützen sie die erweiterten Suchfähigkeiten [FHLW03].
- Taxonomien unterstützen in der Praxis auch den Suchvorgang. Laut Fensel et al. [FHLW03] kann die Ausweitung der Anfrage unter bestimmten Bedingungen radikal das Suchergebnis verbessern. Diese Bedingungen sind zum Beispiel die kurze Dokument-Länge und begrenzte Inhaltsbereiche. Die Methode der Anfragenerweiterung dient somit der Ausweitung der Benutzer-Anfragen mit Begriffen von spezifischeren Kategorien in der Hierarchie [FHLW03].
- Taxonomien bieten Unterstützung bei der Begriffserklärung und der entsprechenden Bedeutung. Wenn das gleiche Wort des Öfteren in verschiedenen Taxonomie-Bereichen auftaucht, wird für die Anwendung unerlässlich die Bedeutung des Wortes zu eruieren. Die Anwendung inkludiert ein allgemeineres Niveau in der Taxonomie. Beinhaltet eine Ontologie die Information, dass *Jordan* eine Instanz von *BasketballPlayer* sowie eine Instanz von einem *Land* ist, dann wird in der Anwendung erforderlich sein zwischen den beiden Anfragen wählen zu können [FHLW03; McGu02].

Ein Beispiel für eine Ontologie-Sprache aus der Praxis stellt die *Simple HTML Ontology Extension* dar, welche in der Literatur auch unter SHOE-Ontologie bekannt ist. Das besondere an SHOE ist, dass sie erlaubt die Inhalte einer Webseite maschinen-verständlich zu gestalten. Sie versucht somit dem Ziel des semantischen Webs gerecht zu werden. Denn durch maschinen-interpretierbare Semantik lässt sich die Zusammenarbeit zwischen Benutzer und Computer optimieren. Im Folgenden werden SHOE-Ontologien vorgestellt und einige Beispiele präsentiert.

SHOE-Ontologien

Simple HTML Ontology Extension (SHOE) ist eine XML-kompatible Sprache zur Wissensrepräsentation im Web. SHOE wurde auf der Universität von Maryland entwickelt, um Informationen über Webseiten und Dokumente auch maschinen-interpretierbar zu gestalten und damit dem Benutzer signifikant bessere Suchresultate zu ermöglichen [Cove00].

SHOE verwendet Ontologien um gültige Elemente, die zur Beschreibung von Entitäten herangezogen werden können, zu definieren. Jede Ontologie kann andere Ontologien wieder verwenden. Eine Ontologie wird in einer HTML oder XML Datei gelagert und wird den Autoren des Dokuments und den SHOE Agents zur Verfügung gestellt. Diese Datei inkludiert Tags, die eine Ontologie identifizieren können, bezeichnen welche Ontologien erweitert wurden und definieren verschiedene Elemente der bestimmten Ontologie [FHLW03]. Ein Beispiel einer SHOE-Ontologie mit dem Namen „Diplomarbeit-ont“ wird im Bild 3 gezeigt.

In der SHOE-Syntax taucht eine Ontologie zwischen dem Tag `<ONTOLOGY ID=id VERSION=version>` und dem Tag `</ONTOLOGY>` und wird anhand der Kombination von *id* und *version* identifiziert. Eine Ontologie kann Kategorien, Relationen und andere Komponenten definieren, indem es zu diesem Zweck spezielle Tags inkludiert [FHLW03].

Eines der Beziehungen, das in dem Ontologie-Beispiel im Bild 3 definiert wird, ist *beratet*. `<DEF-RELATION NAME =“beratet“>` stellt die Beziehung zwischen der Fakultät und dem/der StudentIn dar. Das Argument der Relation kann auch einen *string*, *number*, *date* oder *boolean-Wert* einnehmen, so wie beispielsweise das zweite Argument von der Beziehung *schreibtDiplomarbeit* im Bild 3 [FHLW03]. Das Ontologie-Beispiel definiert auch mehrere Kategorien, so ist das *Institut* eine Subkategorie von *Universität*. Der Tag `<DEF-RELATION>` wird zur Bildung von relationalen Definitionen verwendet.

Das Ontologie-Beispiel im Bild 3 definiert Deduktionsregeln, die durch den `<DEF-INFERENCE>` Tag indiziert werden um zusätzliche Axiome zu unterstützen. Die `<INF-IF>` und `<INF-THEN>` Tags indizieren die Bezugselemente (*Bedingungsteil*) und entsprechende Deduktion (*Rückschlüsse*). Eine SHOE-Deduktionsregel besteht aus Bedingungsteil und Aktionsteil [FHLW03].

Es gibt drei Untergliederungen von Deduktion (*Inference*): Kategorie, Relation und Vergleich. Die Argumente der Untergliederung können konstant oder variabel sein. Letzteres wird mit dem Schlüsselwort `VAR` indiziert [FHLW03]. Das Ontologie-Beispiel im Bild 3 inkludiert eine Regel, die besagt, dass das *Thema* der vorliegenden Arbeit dem Bereich der *Informationswirtschaft* zugeordnet ist. Da der Titel durchaus im Laufe der Erarbeitung einer Diplomarbeit auch abweichen kann, wird sie als variabel definiert. Zudem besteht die Möglichkeit einen Haupttitel und einen Untertitel zu verfassen. Im Zusammenhang mit der vorliegenden Arbeit wurde vom Untertitel nicht Gebrauch gemacht, so dass der Untertitel auch *boolean* definiert werden könnte.

Der Tag `<DEF-CATEGORY>` wird zur Bildung von Definitionen der Kategorien verwendet. Damit werden die Kategorien näher spezifiziert. Folglich kann dadurch zwischen verschiedenen Instanzen klassifiziert werden. Kategorien können als Subkategorien unter Superkategorien gruppiert werden. Bild 3 zeigt unter anderem die Definitionssyntax von Kategorien. Ein Beispiel für eine SHOE-Instanz wird im Bild 4 dargestellt.

```

<!-- Diese Ontologie heißt „Diplomarbeit-ont“. -->
<ONTOLOGY ID="Diplomarbeit-ont" VERSION="1.0">

<!-- Einige Elemente werden aus einer bestehenden Ontologie
herangezogen, mit dem Präfix „a“ -->

<USE ONTOLOGY ID="WuWien-ont" VERSION="1.0" PREFIX="a"
URL=http://www.wu-wien.ac.at/>

<!-- Lokale Pseudonamen für bestimmte Begriffe anlegen -->

<DEF-RENAME FROM - "a.Thema" TO="Thema der Diplomarbeit">
<DEF-RENAME FROM - "a.Autor" TO="Autor der Diplomarbeit">
<DEF-RENAME FROM - "a.Betreuer" TO="Betreuer">

<!-- Kategorien und Beziehungen der Subkategorien definieren -->

<DEF-CATEGORY NAME="Rektor" ISA="Universität">
<DEF-CATEGORY NAME="Institutsmitarbeiter" ISA="Institut">
<DEF-CATEGORY NAME="Abteilungsleiter" ISA="Abteilung">
<DEF-CATEGORY NAME="Name" ISA="Betreuer">
<DEF-CATEGORY NAME="Diplomandin" ISA="Studentin">

<!-- Beziehungen definieren, das Beispiel hier ist binär,
Beziehungen können aber auch n-Grades sein -->

<DEF-RELATION NAME="beratet">
  <DEF-ARG POS="1" TYPE="Institutsmitarbeiter">
  <DEF-ARG POS="2" TYPE="Diplomandin">
</DEF-RELATION>

<DEF-RELATION NAME="schreibtDiplomarbeit">
  <DEF-ARG POS="1" TYPE="StudentIn">
  <DEF-ARG POS="2" TYPE=".BOOLEAN">
</DEF-RELATION>

<!-- Definiton einer Deduktion, die besagt, dass das Thema dieser
Diplomarbeit der Informationswirtschaft zugeordnet ist -->

<DEF-INFERENCE>
  <INF-IF>
    <RELATION NAME="a.Thema">
      <ARG POS="1" VALUE="Haupttitel" USAGE="VAR">
      <ARG POS="2" VALUE="Untertitel" USAGE="VAR">
    </RELATION>
    <CATEGORY NAME="Diplomarbeit" FOR="y" USAGE="VAR">
  </INF-IF>
  <INF-THEN>
    <CATEGORY NAME="Informationswirtschaft" FOR="x"
USAGE="VAR">
  </INF-THEN>
</DEF-INFERENCE>
</ONTOLOGY>

```

Bild 3 Beispiel einer SHOE-Ontologie [FHWL03]

```
<INSTANCE KEY="http://wwwai.wu-wien.ac.at">
<!-- Verwendet die Semantik der Ontologie "Diplomarbeit-ont", und
setzt ein "d." davor -->
    <USE-ONTOLOGY ID="Diplomarbeit-ont" VERSION="1.0" PREFIX="d"
        URL="http://www.wu-wien.ac.at/~h9950128/">
<!-- Beansprucht Kategorien für diese und andere Instanzen. -->
    <CATEGORY NAME="d.Diplomandin">
    <CATEGORY NAME="d.Betreuer" FOR="http://wwwai.wu-wien.ac.at">
<!-- Beansprucht Objekte und Beziehungen. -->
    <RELATION NAME="d.Titel">
        <ARG POS="TO" VALUE="Wirtschaftliche Bedeutung von
            Semantic Web Mining und
            Recommender Systemen">
    </RELATION>
    <RELATION NAME="d.beratet">
        <ARG POS="TO" VALUE="http://wwwai.wu-wien.ac.at">
    </RELATION>
</INSTANCE>
```

Bild 4 Beispiel einer SHOE-Instanz [FHWL03]

Tool

Effektiver und effizienter Einsatz des semantischen Webs erfordert *Tools* um die gesamte Kapazität dieser Technologie auszuschöpfen. Insbesondere sind folgende Elemente diesem Zweck unentbehrlich [FHLW03]:

1. Formale Sprachen zur Repräsentation von Ontologien wie zum Beispiel RDF, DAML-O, UPML
2. Editoren und semiautomatische Darstellungsformen zur Bildung neuer Ontologien
3. Wiederverwendbarkeit und Zusammenschließung von Ontologien anhand von Ontologiumgebungen, die die Schaffung neuer Ontologien durch Wiederverwendbarkeit der

bereits vorhandenen Ontologien fördern. (Von Ontologieumgebungen kann nur gesprochen werden wenn entsprechende Ontologiesprachen zum Einsatz kommen und die Kompatibilität zwischen den Systemen gewährleistet ist. Die bereits vorgestellte Architektur von Tim Berners-Lee zeigt im Detail die erforderlichen Bestandteile einer stabilen Ontologieumgebung.)

4. Annotationstools um unstrukturierte und semistrukturierte Informationsquellen mit Metadaten zu verlinken
5. Tools für den Informationszugriff und für die Navigation, die den intelligenten Informationszugriff für den Benutzer ermöglicht
6. Integration- und Umsetzungsdienste zwischen verschiedenen Ontologien, die den Informationsaustausch ermöglichen

Ein Beispiel dieser Technologie stellt laut Fensel et al. [FHLW03] der Protégé Editor dar. Gemäß der Definition ist Protégé ein open-source Ontologie-Editor und ein wissensbasierendes Bezugssystem. Protégé basiert auf Java, ist erweiterbar und stellt ein Fundament für benutzerspezifische und wissensbasierende Anwendungen dar. Protégé unterstützt Frames, XML Schema, RDF(S) und *Web Ontology Language* (OWL). Es stellt praktisch eine Umgebung dar, anhand derer Prototyping und Anwendungsentwicklung im Protégé rascher realisiert werden können. Im Folgenden wird näher auf den genannten Editor eingegangen.

Ontologie Editor: Protégé

Protégé ist eine kostenlose, open-source Plattform, die einer wachsenden Benutzer-Community eine Reihe von Tools bietet. Insbesondere Tools zum Aufbau benutzerspezifischer Modelle und wissensbasierende Anwendungen mit Ontologien sind hierbei zu nennen. Der Kern des *Protégé Editors* liegt im Aufbau von zahlreichen wissensmodellierten Strukturen (*Knowledge-Modeling Structures*) und Abläufe zur Unterstützung von Erstellung, Visualisierung und Manipulation von Ontologien in verschiedenen Repräsentationsformen. Protégé kann kundenspezifisch manipuliert werden um benutzerfreundliche Unterstützung bei der Erstellung von Wissensmodellen zu gewährleisten. Auch eine Erweiterung des *Protégé Editors* in Form einer *plug-in Architektur* oder eine Java-basierende *API (Application Programming Interface)* kann für die Erstellung von wissensbasierten Tools und Anwendungen herangezogen werden [Stan05].

Laut Literatur [Stan05] stellt die Protégé-Plattform zwei Möglichkeiten der Wissensmodellierung mit Ontologien zur Verfügung:

- Der *Protégé-Frames Editor* ermöglicht Benutzern die Erstellung von frame-basierenden Ontologien gemäß *Open Knowledge Base Connectivity protocol* (OKBC). Laut diesem Modell beinhaltet eine Ontologie *eine Reihe von Klassen*, die in einer Hierarchie organisiert sind, *eine Reihe von Vertiefungen*, die mit Klassen assoziiert sind und die Eigenschaften und Beziehung der Klassen beschreiben, und *eine Reihe von Instanzen dieser Klassen*, das sind individuelle Muster dieser Konzepte, die spezifische Werte für ihre Eigenschaften aufweisen.

- Der *Protégé-OWL-Editor* ermöglicht Benutzern die Erstellung von Ontologien für das semantische Web, insbesondere mit *Web Ontology Language* (OWL) von W3C. Eine OWL Ontologie kann Bezeichnungen von Klassen, Eigenschaften und Instanzen beinhalten. Die formale Semantik von OWL spezifiziert wie logische Folgerungen abgeleitet werden. Fakten, die nicht buchstäblich in der Ontologie gegenwärtig sondern durch Semantik bedingt sind, werden auf Grund der Semantik im Rahmen von Schlussfolgerungen gezogen.

3.2 Information aus dem Web

Der Einsatz von Metadaten und weiteren Ontologiesprachen ist somit von wesentlicher Bedeutung für die Beschaffung von Informationen aus dem Web. Das primäre Ziel der digitalen Wertschöpfung liegt in der Optimierung des Einsatzes der im Kapitel 3.1 vorgestellten Technologien durch welche simultan auch eine optimierte Informationssuche realisiert werden kann. Mit anderen Worten führt die Optimierung des einen Einflussfaktors auch zur Optimierung des Anderen, wenn auch nur im ungleich proportionalen Verhältnis. Das Potential der Informationsgewinnung lässt sich somit von der Optimierung der *Technologie für das Web* ableiten. Dieses Argument spricht für die gewählte Reihenfolge der genannten Einflussfaktoren.

Für den Einflussfaktor *Information aus dem Web* spielen aber auch Marketingaspekte wie Marktverhalten, Marketingforschung, Marketingentscheidungen wesentliche Rolle. Um die unternehmerischen Ziele wirksamer und wirtschaftlicher zu erreichen ist es erforderlich die Bedürfnisse des Zielmarktes zufrieden zustellen. Diesem Zweck dienen die Ermittlung des Zielmarktes und der relevanten Kundengruppen sowie die Entscheidung mit welchen Maßnahmen Käufe herbeigeführt bzw. beeinflusst werden sollen [HKMW01].

Die erforderliche Ermittlung der Bedürfnisse des Zielmarktes wird durch Gewinnung von wesentlichen Informationen aus dem semantischen Web realisiert. Der Markt stellt besondere Anforderungen an das Unternehmen hinsichtlich der erforderlichen Informationsgewinnung über Besucher einer Webseite bzw. potentielle Kunden. Denn die Nachfrage nach einem Produkt oder einer Dienstleistung lässt sich erst zufrieden stellen, wenn sie anhand der Informationsgewinnungsverfahren ermittelt werden. Im Web kann das Verhalten der Besucher einer kommerziellen Webseite analysiert werden, um die dabei beobachtbare Tendenz nach einem bestimmten Produkt festzustellen. Diese Tendenz lässt sich mittels Trend-Analysen als Nachfrage nach einem bestimmten Produkt interpretieren. Daher wird in diesem Kapitel insbesondere auf die Ansätze des *Data Mining* und *Semantic Web Mining* eingegangen.

3.2.1 Data Mining

Folgende Hauptaufgabenbereiche des *Knowledge Discovery in Databases (KDD)* oder *Data Mining Ansätze* werden laut Hippner et al. [HKMW01] genannt:

- *Assoziierung*: Beschreibung von Mustern von zusammenhängend auftretenden Elementen in Daten dient der Layout-Planung von Supermärkten oder der Planung von Werbeaussendungen. Einige Beispiele sind die Untersuchung von Transaktionsdaten in Form von Warenkörben, die Berücksichtigung von Transaktionszeitpunkten und die Reihenfolgen von Versicherungsabschlüssen.
- *Regression und Prognose*: Modellierung von Abhängigkeitsbeziehungen zwischen üblicherweise einer abhängigen Variablen und einer Menge unabhängiger Variablen; die Beschreibung der Beziehung erfolgt anhand eines Parametervektors. Dieser Parametervektor kann für erklärende (z.B. Querschnittsuntersuchungen, die der Aufdeckung von Zusammenhangsstrukturen zwischen den Variablen dienen) und für prognostische Zwecke (Zeitreihen, bei denen unter anderem die abhängigen Variablen als verzögerte Regressoren dienen) eingesetzt werden.
- *Klassifikation*: die Aufgabe besteht darin von einer Menge von p Variablen $\{x_1, x_2, \dots, x_p\}$ auf die Klassifikationszugehörigkeit $c \in \{1, \dots, C\}$ eines Objekts zu schließen.
- *Segmentierung*: Dabei werden Kunden auf der Basis von sozioökonomischen Merkmalen, Verhaltensmustern und Einstellungsmerkmalen segmentiert, um die Untergruppen damit effizienter analysieren zu können und folglich auch gezielter ansprechen zu können. In diesem Zusammenhang findet häufig die Anwendung von Clusteranalysen statt.
- *Abweichungserkennung*: Identifikation und Beschreibung von unsystematisch abweichenden Beobachtungen (Ausreißer) oder Beobachtungen, die sich systematisch von anderen unterscheiden. Bei der Regression und Prognose können Ausreißer zu verzerrten Schätzungen führen und daher ist es besonders wichtig diese zu identifizieren und auszuselektieren. Die Methoden der Segmentierung dagegen dienen vielmehr der Ausreißeridentifikation.

Aus betriebswirtschaftlicher Sicht werden aus verschiedenen Alternativen jene ausgewählt, die den größten Erfolg versprechen [HKMW01].

Mehrfach in der Literatur werden *Data Mining* und *Knowledge Discovery in Databases* (KDD) synonym verwendet. Zwecks Unterscheidung wird im Folgenden *Knowledge Discovery in Databases* als *Data Mining im weiteren Sinn (Data Mining i.w.S)* und die Phase des Data Mining in KDD als *Data Mining im engeren Sinn (Data Mining i.e.S)* aufgefasst.

Data Mining i.w.S

“Data Mining, in its most general sense, is about extracting patterns from this mass of data to make business more productive.” [LiBe01]

Die Datenmustererkennung, der deutsche Begriff für *Data Mining*, wird laut Hippner et al. [HKMW01] anhand von zwei Eigenschaften charakterisiert:

- Die Datenmustererkennung arbeitet mit großen, strukturierten Beständen numerischer, ordinal- oder nominalskaliertes Daten, in denen interessante, aber schwer auffindbare Zusammenhänge vermutet werden können.
- Die Datenmustererkennung dient dem Ziel anhand allgemein verwendbarer, effizienter Methoden die bedeutsamsten und aussagekräftigsten Muster zu identifizieren um sie anschließend dem Anwender als interessantes Wissen zu präsentieren.

Typische Anwendungsfelder des *Data Mining* in Vertrieb und Marketing sind Preisfindung, Database Marketing, Marktsegmentierung, Risikomanagement, Kundenbindung und Warenkorbanalyse [ScBM99].

Kurz zusammengefasst besteht ein *Data Mining Prozess* laut Hippner et al. [HKMW01] aus folgenden Phasen:

- Definition des Anwendungsproblems (Ziel)
- Datenbeschaffung und Datenverständnis
- Datenaufbereitung (Analyse und Vorverarbeitung der Daten: Produktnutzungsverhalten, Soziodemographische Merkmale, Kodierung)
- Modellierung (Auswahl der Modellierungstechnik, Testplan, Parameter der Modellierung)
- Bewertung der Modellierung
- Umsetzung/Anwendung

Eine Möglichkeit Daten zu strukturieren und zu analysieren bietet die Methode der Clusteranalyse. Die Methode wird im Folgenden kurz beschrieben.

Clusteranalyse

Das Ziel einer Clusteranalyse ist die Zusammenfassung einer Menge von Klassifikationsobjekten (zum Beispiel Produkte oder Kunden) in homogene Gruppen unter der Einhaltung möglichst hoher Heterogenität zwischen den einzelnen Clustern [HKMW01].

Die anschauliche und problemorientierte graphische Aufbereitung anhand von Clustern dient der Interpretation und Darstellung der Empfehlung zur weiteren Behandlung der entsprechenden Kundensegmente [HKMW01].

Anwendung der Clusteranalyse

Für jede Webseite auf der zugleich mehrere Dienste angeboten werden, kommt es auf die sinnvolle Gestaltung an. Zudem wird die Unterscheidung der verschiedenen Benutzer erforderlich um die Nutzung einzelner Dienste sinnvoll bewerten zu können [Spil01].

Die Webseite Mobilkom-Austria bietet unter anderem folgende Dienste an:

- Handy, Handyzubehör
- Service um Tarife, Online-Rechnung etc.
- Vodafone (MMS, Videotelefonie, E-Mail etc.)
- Nationale Auskunft
- Internationale Auskunft

Ein Benutzer kann vom Auskunftsdienst Gebrauch machen um eine nationale oder internationale Telefonnummer ausfindig zu machen. Ein anderer Benutzer kann zugleich in seine Online-Rechnung der letzten Monate Einsicht nehmen. Ein dritter Benutzer hat die Möglichkeit nach Wahl seines Wunsch-Handys und unter der Voraussetzung vorgegebener Benutzermerkmale ein neues Handy anzufordern. Diese drei genannten Dienste werden von verschiedenen Benutzern in Anspruch genommen. Daher ist es für die Erfolgskontrolle jedes einzelnen Dienstes erforderlich zwischen den Benutzern zu unterscheiden. Die Gruppierung der verschiedenen Benutzer dient der sinnvollen Unterscheidung der angebotenen Dienste einer Webseite sowie der adäquaten Anwendung der Erfolgskontrolle. Jeder einzelne Dienst wird anhand jener Benutzer analysiert, die den Service auch in Anspruch nehmen. Hierbei wird das Verhalten und die Zufriedenheit der Benutzer analysiert bzw. gemessen. Laut Spiliopoulou [Spil01] wird zwischen folgenden zwei Benutzergruppen unterschieden: *aktiver Benutzer* und *Kunde*. In der Literatur wird der *aktive Benutzer* als ein Besucher einer Webseite definiert, der von einem Rechner auf diese Webseite zugreift. Der Zugriff auf eine Webseite ist dabei entscheidend. Dieser Zugriff kann aus unterschiedlicher Motivation heraus erfolgen und ist von den Bedürfnissen des aktiven Benutzers abhängig [Spil01]. Bild 5 zeigt drei verschiedene Dienste welche durch die Art des Zugriffs auf die Webseite charakterisiert sind.

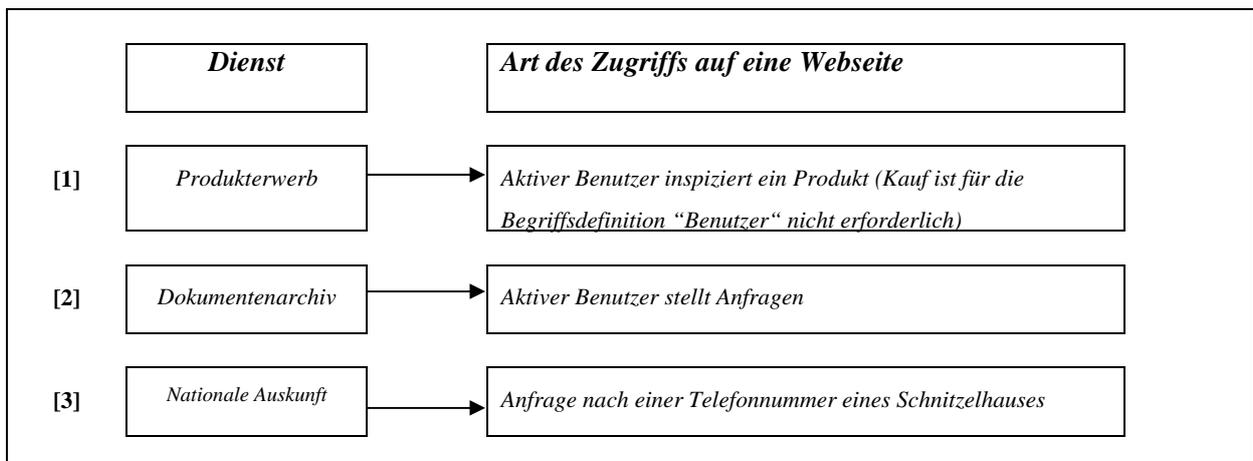


Bild 5 Art des Zugriffs auf eine Webseite [Spil01]

Ein Besucher einer Webseite ist laut Literatur erst dann ein Kunde, wenn er auch die objektiven Ziele der einzelnen Dienste erfüllt hat. Zum Beispiel ist beim Erwerb von Gütern das objektive Ziel der Kauf eines Produkts [Spil01]. Ob die Besucher tatsächlich auch potentielle Kunden werden, hängt sowohl von der Struktur als auch vom Angebot der Webseite ab. Sowohl äußere als auch innere Faktoren einer Webseite beeinflussen den Kunden in seiner Entscheidungsfindung. Zu den äußeren Faktoren lassen sich oberflächliche Strukturen und Design aber auch die Einstellung des Besuchers selbst nennen. Zu den inneren Faktoren zählen vor allem die *inneren Werte* also qualitative Aspekte, wie die Qualität des Angebots oder den Kunden ansprechenden *Preis-Mengen* bzw. *Preis-Qualitäts-Verhältnis* [LiBe01].

Data Mining i.e.S

“The purpose of data mining is not merely to analyze data – the purpose is to add value to business.” [LiBe01]

Laut Linoff et al. [LiBe01] liegt der Zweck des *Data Mining* nicht bloß in der Analyse der Daten sondern insbesondere in der zusätzlichen Wertschöpfung. Daher ist das Hauptmerkmal des *Data Mining i.e.S* der zusätzliche Mehrwert, der durch Einsatz von Data Mining Methoden realisiert wird. Gerade aufgrund dieses Merkmales wurde die Unterscheidung zwischen *Data Mining i.w.S* und *Data Mining i.e.S* vorgenommen.

Laut Bloehdorn [Bloe01] beschreibt der Begriff *Data Mining (Data Mining i.e.S)* jenen Teil des *Knowledge Data Discovery Processes* (KDD-Prozess), der sich der eigentlichen Mustererkennung widmet. Der gesamte KDD-Prozess setzt sich zusammen aus der Aufbereitung und Vorbereitung der Daten (*Preprocessing*), der Mustererkennung und Ergebnisinterpretation [Bloe01]. Bedeutender Bestandteil des Data Mining Prozesses ist das Verständnis von Kundenmustern, die in Web-Daten zu finden sind.

Im Folgenden wird auf die Aufbereitung und Vorbereitung der Daten eingegangen. Dabei werden laut Literatur drei Phasen unterschieden: *Datenreinigung*, *Trennung der Aktivitäten unterschiedlicher Benutzer* und die *Gruppierung in Sitzungen*.

1. Datenreinigung:

Das Ziel der Analyse ist es zu bestimmen welche Einträge nicht relevant sind. Zur Analyse der Webseiten-Nutzung durch potentielle Kunden werden spezifische Einträge herangezogen, nämlich jene der menschlichen Benutzer. Die restlichen Einträge sind irrelevant und werden während der Datenreinigung entfernt. Die restlichen Einträge können beispielsweise Tätigkeiten von Robotern und anderen vorprogrammierten Werkzeugen umfassen [Spil01]. Roboter agieren unabhängig und werden vom menschlichen Benutzer nicht gesteuert. Der Einsatzbereich solcher Roboter erstreckt sich von der Indizierung von Webseiten für Suchmaschinen bis zur Filterung von Produkt- und Preisinformationen für beispielsweise automatische Preisvergleiche [Gasc02, HiMW02].

2. Trennung der Aktivitäten unterschiedlicher Benutzer:

Zuerst wird zwischen Nutzern unterschieden, die von demselben Host zugreifen [Spil01]. In EDV-Räumen der Wirtschaftsuniversität Wien wird jeder Rechner von verschiedenen Benutzern tagtäglich verwendet. Unabhängig von der Matrikelnummer des Studierenden wird für jeden Zugriff, die Adresse des Host des zugreifenden Rechners protokolliert.

In vielen Unternehmen finden Zugriffe aufs *World Wide Web* von allen Mitarbeitern durch einen einzigen Rechner (*Proxy Server*) statt. Die einzelnen Rechner der Mitarbeiter werden auf den Proxy Server umgeleitet um höhere Sicherheit und Geschwindigkeit zu gewährleisten [Spil01]. Somit werden alle Mitarbeiter des Unternehmens, die durch den Proxy Server zugreifen als *eine* Person identifiziert. Die Verwendung von *Cookies* auf Webseiten erlaubt die Identifizierung jedes Benutzers bei einmaligen oder mehrmaligen Zugriffen auf die entsprechende Webseite [Spil01].

Technisch ist es möglich und in manchen Ländern (zum Beispiel in Deutschland das Teledienststedatenschutzgesetz) sogar verpflichtend, Cookies und ähnliche skriptbasierte Mechanismen so einzusetzen, dass keine personenbezogenen Daten aufbewahrt werden [Spil01]. Eine Alternative zu Cookies stellt die Methode der unterschiedlichen HTTP-Clients dar.

Cooley et al. [CoMS99] gehen davon aus, dass Zugriffe, welche unterschiedliche HTTP-Clients verwenden, von unterschiedlichen Benutzern aufgerufen wurden (auch wenn sie vom selben Rechner kommen).

3. Gruppierung in Sitzungen

Hierbei werden zusammengehörende Zugriffe zu einer Sitzung zusammengefasst. Wie lange nun eine Sitzung dauert bzw. welche Zugriffe zusammengehören, hängt vom Anwendungsbereich und vom Thema der zugegriffenen Webseiten ab [Spil01]. Bild 6 zeigt drei verschiedene Möglichkeiten des Zugriffs auf die Webseite der Mobilkom Austria (<http://www.a1.net>).

Beim ersten Zugriff wünscht der Benutzer einen Tariffwechsel [1]. Aus dieser Motivation heraus recherchiert er auf der Webseite des Anbieters. Nach einer ca. halbstündigen Sitzung entscheidet sich der Benutzer für eine Vertragsänderung. Beim zweiten Zugriff benötigt der Benutzer lediglich eine Information über seine aktuelle Rechnung [2], daher dauert diese Sitzung lediglich fünf Minuten. Beim letzten Zugriff fordert der Benutzer ein neues, bereits ausgesuchtes Handy an [3] und benötigt daher nicht länger als zehn Minuten.

Webseite	Thema	Anwendung	Resultat	Dauer der Sitzung
[1] a1.net	Tariffwechsel	Recherche	Vertragsänderung	ca. 30 min
[2] a1.net	Online Rechnung	spezifischer Bereich	Information	ca. 5 min
[3] a1.net	Neues Handy	Suche im Angebot	Handy anfordern	ca. 10 min

Bild 6 Beispiele für verschiedene Web-Zugriffe

Laut Linoff et al. [LiBe01] werden unter anderem folgende Daten der Benutzer einer Webseite gespeichert und für weitere Analysen herangezogen:

- Häufigkeit des Onlinekaufs
- Zeit und Datum des Onlineverkaufs
- durchschnittliche Größe des Warenkorbs
- Gesamtanzahl der verkauften Produkte
- Gesamtanzahl der ausverkauften Produktkategorien
- Reaktionen auf Empfehlungen und auf Onlineprämien

Die relevanten Informationen geben in weiterer Folge Aufschluss darüber, zu welcher Zeit interessierte und potentielle Kunden einen Kauf tätigen, zu welcher Jahreszeit oder in welchem Monat die meisten Käufe getätigt werden, welche Reaktionen die angebotenen Prämien verursachen, welche Prämien am beliebtesten sind, welche Produkte am häufigsten gekauft werden, welche Produkte wegen zu langer Wartezeit wieder abbestellt wurden usw.

Intelligente Informationsagenten

Klusch [Klus01] definiert den Informationsagenten indem er die notwendigen Anforderungen wie autonomes, zielgerichtetes Verhalten, soziale Kooperation, Intelligenz und Lernfähigkeit in die Begriffsdefinition integriert.

Die Anforderungen an solche Agenten sind im Folgenden spezifiziert. Laut Klusch [Klus01] integriert der Begriff des Informationsagenten die Erfüllung zumindest einer der im Folgenden angeführten Anforderungen:

- Erwerb und Management von Informationen
- Synthese und Präsentation von Informationen
- intelligente, individuelle Benutzerassistenz

In der zugrunde liegenden Literatur [Klus01] erfolgt folgende Klassifikation von Informationsagenten:

- Kooperative Informationsagenten und Systeme
- Adaptive Informationsagenten

- Rationale Informationsagenten für den elektronischen Handel
- Mobile Informationsagenten

Im Folgenden werden die *mobilen Informationsagenten* näher beschrieben, da diese Agenten in Softwareapplikationen für insbesondere mobile Telekommunikationsgeräte eingesetzt werden. Daher steigt auch die Bedeutung ihres Einsatzes in der Telekommunikationsbranche bzw. im Bereich des Mobilfunknetzes, welches zum Gegenstand des analytischen Teils dieser Arbeit (Kapitel 4) gewählt wurde.

Mobile Informationsagenten

In Bezug auf ein ort- und zeitunabhängiges, intelligentes Informationsmanagement steht die Mobilität von Informationsagenten im Vordergrund. Zudem sind Mobile Agenten in der Lage unter gegebenen Umständen für die Ausführung ihrer Aufgaben auf mehrere, geographisch verteilte Server zu migrieren [Klus01].

Die Verwendung solcher Agenten in geeigneten Softwareapplikationen wird durch die steigende Verbreitung von mobilen Telekommunikationsgeräten unterstützt. Zur Begriffsdefinition des mobilen Telekommunikationsgerätes ist die Verbindung zum Internet, zum Breitband und zum drahtlosen, satellitengestützten Datennetzwerken ausschlaggebend. Beispiele dafür sind Mobiltelefone, die mit *Wireless Application Protocol* (WAP) ausgestattet sind, Kommunikatoren (wie zum Beispiel der *Vocera Communicator*: <http://www.vocera.com/>) und *Personal Digital Assistent* (PDA) [Klus01]. Der Kommunikator, der vom kalifornischen Unternehmen *Vocera* entwickelt wurde, ist ein Kommunikationssystem (http://www.vocera.com/downloads/voc_ds_v5_091405_US.pdf), der *Push-to-talk Gespräche* und *Spracherkennung* ermöglicht. Wobei sich die Benutzer innerhalb einer bestimmten Reichweite befinden müssen. In der Praxis findet der Einsatz oft innerhalb eines Krankenhauses statt.

Das Hauptaufgabengebiet für mobile Agenten betrifft dezentrale, effiziente Dienstleistungen in der Telekommunikation [WeVe98]. In diesem Kontext werden auch Sicherheitsaspekte diskutiert, da allein schon durch die Internet-Verbindung Viren auf mobile Geräte übertragen werden können. Laut der zugrunde liegenden Literatur werden Lösungsansätze vorgeschlagen. So können unter anderem vertrauenswürdige Ausführungsumgebungen für mobile Agenten auf fremden Server für höhere Sicherheit sorgen. Zudem können Mechanismen zur Entdeckung bzw. Vorbeugung von Codemodifikationen integriert werden wie zum Beispiel durch Verwendung sicherer Hardware oder durch kryptografische, zeitbegrenzte Ausführungen von Agenten [Klus01].

3.2.2 Semantic Web Mining

Laut Berendt et al. [BeHS02] setzt sich *Semantic Web Mining* aus zwei bedeutenden Forschungsgebieten zusammen:

- *Semantic Web* und
- *Web Mining*

Die Idee des *Semantic Web Mining* ist einerseits die Verbesserung des *Web Mining* anhand von semantischen Strukturen im Web, und andererseits die Verwendung von *Web Mining* zum Aufbau des semantischen Webs [BeHS02].

Semantic Web

Semantic Web ist die von Tim Berners-Lee vorgestellte Idee, eine Vision zur Unterstützung des Benutzers bei seinen Aufgaben durch maschinen-verarbeitbare Information (*machine-processable information*) [BeHS02]. Das Konzept des Semantic Web wurde bereits in Kapitel 3.1.1 vorgestellt.

Bei *Semantic Web Mining* kommen die Ontologiesprachen wie zum Beispiel RDF, DAML oder OWL im Zusammenhang mit den Methoden des Web Mining zur Anwendung. Die drei Methoden des Web Mining, *Web Content Mining*, *Web Structure Mining* und *Web Usage Mining* werden anschließend beschrieben um einen kurzen Einblick in den Bereich zu liefern.

Web Mining

Typische Bewegungspfade der Web-Konsumenten lassen sich anhand von Web Mining Methoden identifizieren. In weiterer Folge können häufige Kaufmuster ermittelt werden. Zudem können Bewegungspfade von Web-Konsumenten segmentiert und nach ihrer Kaufwahrscheinlichkeit bewertet werden [HiMW02].

In erster Linie stehen die aufgezeichneten Seitenaufrufe der Besucher in den Logfiles der Webserver zur Verfügung, aus denen sich anhand geeigneter Analyseverfahren die Verhaltensweisen der Webbenutzer und der Erfolg der Sitzungen ableiten lassen [HiMW02].

Die Erstellung deskriptiver Statistiken (Logfile-Analysen) dient der Auswertung von Daten der Web-Benutzer. Der Mangel solcher Logfile-Analysen liegt darin, dass sie nicht in der Lage sind, selbständig Muster in den Daten der Benutzer aufzufinden [HiMW02]. Daher werden automatische Mustererkennungsverfahren genutzt, das heißt *Verfahren des Data Mining* [HiMW02].

Web Mining untersucht drei wesentliche Gegenstände, woraus sich die drei Methoden des Web Mining ableiten lassen:

- die Inhalte (*Web Content Mining*)
- die Struktur der Verlinkung (*Web Structure Mining*)
- das Benutzerverhalten (*Web Usage Mining*)

Web Content Mining ist der Prozess der Gewinnung von nützlicher Information aus dem Inhalt der Web-Dokumente [SrDK05]. Dies geschieht beispielsweise mit Verfahren des Information Retrievals [Wiki05].

Web Structure Mining ist der Prozess, bei dem Informationen aus der Web-Struktur extrahiert werden [SrDK05]. Diese Informationen können zum Beispiel mit Verfahren der Webometrie (auch Cybermetrie, Internetometrie oder Web Metrics genannt) extrahiert werden. Die Webometrie untersucht das Web und das Verhalten der Web-Benutzer [Wiki05].

Web Usage Mining ist der Prozess der Extrahierung von Informationen über Gebrauchsmuster von Links [SrDK05]. Das Benutzerverhalten kann beispielsweise durch die Analyse von Logfiles untersucht werden. Eine Logdatei stellt das Protokoll aller oder bestimmter Aktionen vom Benutzer an einem Rechner, ohne dass die Benutzer in ihrer Arbeit gestört oder sonst wie beeinflusst werden [Wiki05]. *Web Usage Mining* beschreibt somit die Anwendung von Data Mining Techniken zur Auffindung von interessanten Gebrauchsmustern (*usage patterns*) aus Web-Daten. *Usage Data* beinhaltet die Identität oder Herkunft der Web-Benutzer sowie deren Browsing-Verhalten auf einer Webseite [SrDK03].

Mit *Web Usage Mining* wird das Wissen durch die Analyse von Nutzungsdaten ermittelt. Beispiele für die genannten Nutzungsdaten sind Assoziationen zwischen Seiten oder Produkten, Segmentierung des Web Marktes, Eigenschaften von Nutzergruppen sowie die Überprüfung der Qualität und Akzeptanz einer Webseite [Spil01; HKMW01].

Die Beziehungen zwischen den einzelnen Methoden (*Web Content Mining*, *Web Usage Mining*, *Web Structure Mining*, *Impersonalized Web Usage Mining*, *Personalized Web Usage Mining*) lassen sich in hierarchische und horizontale Beziehungen unterteilen. Letzteres deutet darauf hin, dass die Ergebnisse einer Methode auch für die Analyse einer anderen Methode genutzt werden können. So können die Ergebnisse aus *Personalized Web Usage Mining* auch für *Impersonalized Web Usage Mining* angesichts der Marktsegmentierung und Verhaltensmuster von Nutzergruppen angewendet werden [Spil01; HKMW01]. Horizontale Beziehung zwischen zwei Methoden wird im Bild 7 mit der strichlierten Linie gekennzeichnet.

Bild 7 zeigt somit, dass die Ergebnisse des *Web Content Mining* auch für *Web Usage Mining* genutzt werden können und vice versa. Dasselbe gilt auch für *Impersonalized* und *Personalized Web Usage Mining*.

Bild 7 stellt die genannten Methoden des Web Mining im Überblick dar.

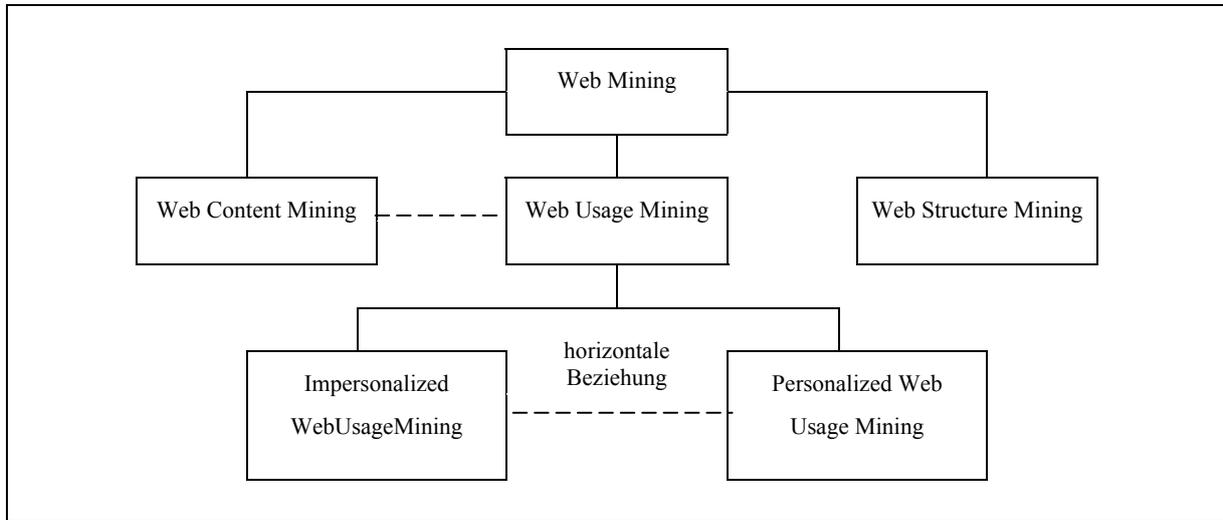


Bild 7 Methoden des Web Mining [HKMW01]

Laut Spiliopoulou [Spil01] werden beim Zugriff auf eine Webseite in der Log-Datei des Web-Servers beispielsweise folgende Daten protokolliert:

- der Host (der Rechner von dem aus der Zugriff erfolgte)
- Datum und Uhrzeit des Zugriffs
- der Uniform Resource Locator (URL) der aufgerufenen Webseite
- Code, der den Erfolg oder Misserfolg des Zugriffs beschreibt
- die Anzahl der gelesenen Bytes uvm.

Bevor aus diesem Datenbestand sinnvolle Informationen herausgefiltert werden können, ist ein Aufbereitungsprozess erforderlich. Dieser Prozess umfasst laut Spiliopoulou [Spil01] die Datenreinigung, die Trennung der Aktivitäten unterschiedlicher Nutzer und ihre Gruppierungen in Sitzungen. Diese Phasen wurden in Kapitel 3.2.1 erläutert.

Die Auswertung des Such- und Einkaufsverhaltens der Benutzer von kommerziellen Webseiten stellt die Grundlage für personalisierte Webseiten im Web dar. Das Wissen über den Kunden und seine Bedürfnisse stellen den entscheidenden Faktor für den erfolgreichen Einsatz von Personalisierungsstrategien dar. Der Einsatz eines Empfehlungssystems macht daher erst dann Sinn, wenn das online-tätige Unternehmen über die relevanten Daten und Informationen wie zum Beispiel über Kundenprofile verfügt.

Konzept des Global Research

Das *Konzept des Global Research* kennzeichnet sich dadurch, dass eine Abfrage an beliebige Datenbanken *gleichzeitig* gestellt werden kann. Im Folgenden wird das *Konzept des Global Research* anhand von *Liman Scout* präsentiert. *Liman Scout* stellt durch sein Webangebot ein Praxisbeispiel für intelligente Informationssuche im Web dar.

Praxisbeispiel: Liman Scout

Ein Beispiel für ein modernes Tool zur Abfrage von Datenbanken im Web stellt das *Liman Scout* dar [Lima05]. Insbesondere kann *Liman Scout* eine Abfrage an nahezu beliebig viele Datenbanken gleichzeitig versenden und die Ergebnisse *unter einer einheitlichen Oberfläche* darstellen. Die einheitliche Abfrage wird mithilfe des Kommunikationsprotokolls Z39.50 realisiert. Dabei wird die eigene Abfrage nach Z39.50 übersetzt und ist somit für alle Datenbanken, die dasselbe Protokoll verwenden, verständlich und mit ihnen kompatibel. Das Protokoll des *American National Standards Institute* (ANSI-Protokoll), Z39.50, ist im Bereich der Datenbanken ein sog. Interpreter, mit dem eine gemeinsame Sprache für eigene und fremde Systeme realisiert wird. Dabei ist es erforderlich, dass jedes System seine eigene Sprache nach Z39.50 übersetzt um von anderen Systemen verstanden zu werden [Brau05].

Die Anwendung des Z39.50 wird insbesondere in Bibliotheken realisiert. Da in den USA bereits eine Vielzahl von Bibliotheken ihre Kataloge über Z39.50 bereitgestellt haben, setzte sich dieser Standard zunehmend auch in Europa durch. Die fünf größten Bibliotheken Deutschlands verfügen bereits über einen Z39.50 Zugang. Durch dieses Angebot ergeben sich für den Anwender völlig neue Möglichkeiten der Literaturrecherche. Weitere Informationsdienste sind auch elektronische Dokumente, Bilder oder beliebige Datenbankinhalte [Lima05].

Anstatt sich bei der Recherche im Web mit jeder einzelnen Datenbank zu konfrontieren, bietet der Lösungsansatz des *global research* eine sinnvolle Alternative. Das Konzept wird durch die Möglichkeit – der *globalen* Abfrage an viele Datenbanken *gleichzeitig* – realisiert [Lima05].

Die erhaltenen Abfrageergebnisse werden von *Liman Scout* lokal verwaltet, das heißt die Einschränkung oder das Sortieren von Suchergebnissen wird von *Liman Scout* durchgeführt, so dass der abzufragende Server lediglich die grundlegenden Dienste des Z39.50-Protokolls unterstützen muss. Daher eignet sich der *Liman Scout* als *Offline-Reader* für die Datenbestände [Lima05].

3.3 Personalisierung im Web

Zum Gegenstand dieses Kapitels wird die Unternehmensphilosophie des *Customer Relationship Managements* (CRM) herangezogen. Die effektive Geschäftsphilosophie des CRM liefert begründeten Anlass für die *Personalisierung im Web*, wodurch sich die Kundenbeziehung pflegen lässt. Durch gezielten Einsatz von Personalisierungssoftware lässt sich der Zufriedenheitsgrad der Kunden erhöhen. Daher wird in diesem Kapitel auch das Recommender System als ein Beispiel für die Personalisierungssoftware vorgestellt.

3.3.1 CRM vs. eCRM

Unter *Customer Relationship Management (CRM)* wird im Allgemeinen eine effektive Geschäftsphilosophie verstanden. Ursprünglich war es ein integriertes Marketing-Instrument, das dazu diente die Zufriedenheit der Kunden zu steigern, indem eine stabile Beziehung zum Kunden aufgebaut wurde. An dem Grundgedanken hat sich für die Geschäftsphilosophie nichts geändert. Vielmehr hat sich der Bereich durch die Kundenidentifizierung, Kundenbestandssicherung und die Optimierung des Kundenwertes zu einer professionellen Unternehmensphilosophie entwickelt.

Der Zufriedenheitsgrad der Kunden stellt für das Unternehmen wesentliches Erfolgspotential dar. *Customer Relationship Management (CRM)* hat zum Ziel den Zufriedenheitsgrad des Kunden zu maximieren. Der Aufbau einer Kundenbeziehung und die damit verbundene Informationssammlung und Informationsaufbereitung ist wesentliche Aufgabe des *Customer Relationship Managements* und dient langfristig dem Ziel der Vertrauensgewinnung [Scha03]. Aus der Sicht des Unternehmens kann durch das Vertrauen und die Treue des Kunden mehrfacher Zusatznutzen erzielt werden.

Liegt der Fokus des CRM im *Electronic Commerce*, dann spricht man von *eCRM*. Das heißt Letzteres ist lediglich ein Teilbereich der CRM-Philosophie. Hauptsächlich ist eCRM dadurch charakterisiert, dass die Interaktion zwischen Unternehmen und Kunden automatisiert und direkt stattfindet, wie es zum Beispiel im Web möglich ist. Die CRM-Software unterstützt lediglich die Mitarbeiter eines Unternehmens in den Bereichen Marketing, Service und Vertrieb, während eine eCRM-Lösung die direkte Interaktion mit dem Kunden ermöglicht, indem der Kunde die Möglichkeit zum direkten Zugriff auf die Angebote des Unternehmens erhält. In diesem Prozess ist kein Mitarbeiter des Unternehmens direkt involviert, daher spricht man in diesem Zusammenhang auch von Selbstbedienung (*Self-Service*). Ein Beispiel solcher eCRM-Lösungen sind kommerzielle Webseiten mit integrierten Online-shops [GVSR01]. Im Zusammenhang mit dem Zugriff auf Informationen im Web über mobile Endgeräte wie Handys und PDAs wird von mobilem CRM (*mCRM*) gesprochen [GVSR01].

Clemens [Clem03] differenziert zwischen verschiedenen Möglichkeiten der Mehrwertgenerierung aus der Sicht des Benutzers. Der Mehrwert lässt sich somit laut Clemens [Clem03] im Zusammenhang mit personalisierter, vom Empfänger explizit gewünschter, Leistungen wie folgt differenzieren:

- Materieller Mehrwert
- Mehrwert durch Informationsnutzen

- Mehrwert durch Unterhaltungsnutzen

Beispiele solcher Leistungen können Anfragen hinsichtlich Wetter des Tages, wirtschaftliche Neuigkeiten oder etwaige Produktinformationen sein. Personalisierte Mehrwert-Angebote stiften dem Unternehmen insofern einen Nutzen, inwieweit sie auch ihre Funktion erfüllen. Mehrwert-Angebote können ihre Funktion unterschiedlich gut erfüllen, was wiederum auch von der Einstellung zum Angebot sowie der Wahrnehmung des Endbenutzers abhängt.

Um Mehrwert-Angebote entsprechend gut generieren zu können, sind Informationen über den Kunden notwendig, wodurch sich in weiterer Folge Kundenprofile erstellen lassen. Kundenprofile sind unter expliziter Einwilligung des Benutzers als Grundlage für die Generierung und Entwicklung entsprechender, personalisierter Mehrwert-Angebote sowie für den Aufbau einer nachhaltigen Kundenbeziehung heranzuziehen [Clem03].

Laut Linoff et al. [LiBe01] kommt der Differenzierung von verschiedenen Stufen der Kundenbeziehung besondere Bedeutung zu. Diese Stufen lassen sich laut Literatur in *potentielle Kunden* gliedern, die

- zwar die Webseite besucht haben jedoch für den Kauf kein Interesse zeigten [Stufe 1]
- Interesse zeigten jedoch aufgrund von diversen Faktoren keinen Kauf tätigten [Stufe 2]
- zwar kaufen wollten aber den Warenkorb wieder leeren mussten [Stufe 3]

und in *Kunden*, die

- einmal gekauft haben oder [Stufe 4]
- mehrmals gekauft haben [Stufe 5]

Jene Besucher, die auf der Webseite keine interessanten Produkte finden, kommen somit nur bis zur ersten Stufe der Kundenbeziehung des Unternehmens, das die entsprechende Webseite betreibt [Stufe 1]. Auf einer anderen Webseite können diese Besucher aufgrund von inneren oder äußeren Faktoren leider keinen Kauf tätigen und kommen somit nur bis zur zweiten Stufe [Stufe 2]. Die dritte Stufe der Kundenbeziehung betrifft alle Besucher, die zum Kauf entschlossen sind, jedoch aufgrund von technischen Schwierigkeiten der Webseite leider zu keinem Kauf kommen [Stufe 3]. Diese potentiellen Käufer würden bis zur vierten Stufe gelangen, wenn sie den Onlinekauf nach der erfolgreichen Wartung oder Reparatur des Webshops erneut tätigen [Stufe 4]. Insbesondere treue bzw. zufriedene Kunden gelangen bis zur höchsten Stufe [Stufe 5].

Durch die Erstellung von Kundenprofilen lassen sich Informationen für den einzelnen Kunden auch spezifizieren und personalisieren. Die Webseite der *Mobilkom Austria AG & Co KG* (www.a1.net) oder *One GmbH* (www.one.at) haben für den Kunden eigenen spezifischen Bereich – Mobilkom Austria unter *Meine Anwendungen* und One GmbH unter *Kundenservice*. Während auf der One-Webseite *Kundenservice* als eigene Dimension dargestellt wird, ist auf der Mobilkom Austria Webseite der Bereich *Meine Anwendungen* unter der Dimension *Privat* gegliedert. Die Benutzerfreundlichkeit wird somit auf der One-Webseite durch die leichte Auffindbarkeit des benutzerspezifischen Bereichs gewährleistet. Auf der Mobilkom Austria Webseite werden zum Zweck der Benutzerfreundlichkeit Nebennavigationsleisten eingesetzt. Im Kapitel 4.2.2 wird auf die Dimensionen der A1 Webseite bzw. auf die Dimensionen der One Webseite näher eingegangen.

Laut Literatur [EiVa03; MoCS00] lässt sich der Prozess der Web-Personalisierung in folgende Punkte untergliedern:

- Kategorisierung der gesammelten Daten [EiVa03; MoCS00].
- Modellierung hinsichtlich der Beziehung zwischen verschiedenen Datensätzen [EiVa03; MoCS00].
- Analysierung der gesammelten Daten mittels *Content-based Filtering*, *Collaborative Filtering*, *Rule-based Filtering*, und *Web Usage Mining* [EiVa03; MoCS00].
- Festlegung von Produkten bzw. Dienstleistungen, welche durch ein Personalisierungssystem empfohlen werden sollen [MoCS00].

Die Begriffe des *Content-based Filtering*, *Collaborative Filtering* und *Rule-based Filtering* werden im Zusammenhang mit den Recommender Systemen, in Kapitel 3.3.2 behandelt. Vorwegzunehmen ist lediglich, dass Content-based und Collaborative Filtering zu den individualisierten Empfehlungssystemen zählen und dass dabei bestimmte Datensätze hinsichtlich ihrer Ähnlichkeit verglichen werden.

Personalisiertes Portal

Die Basis eines Portals stellen unterschiedlichste unternehmensexterne und –interne Contentquellen dar, wie beispielsweise Dokumente und Contentmanagement Systeme, aktuelle Nachrichten, Aufzeichnungen über den Verlauf der Diplomarbeit inkl. Updates auf der Homepage der Diplomandin etc. Inzwischen werden von den meisten Datenbankherstellern entsprechende Portalframeworks angeboten, die eine performante Integration dieser unterschiedlichen Contentquellen sowie einen sicheren und transparenten Zugriff gewährleisten. In solchen Frameworks können beispielsweise Datenbanktechniken wie mehrstufiges Catching oder indexbasierte Suche zum Einsatz kommen. Entsprechende Protokolle wie *Lightweight Directory Access Protocol* (LDAP) helfen bei der Einbettung in heterogene Umgebungen [Scha03].

Das Protokoll LDAP dient der Kommunikation zwischen Client und X.500-Verzeichnisdienst. X.500 ist ein Verzeichnis für E-Mail-Adressen, Postadressen und andere Informationen. Mittels einer entsprechenden Metadatenmodellierung, können die einzelnen Contentkategorien formalisiert dargestellt werden. Als Beispiel solcher Metadatenmodellierung kann *Dublin Core Metadata Initiative* (DCMI) genannt werden. Der eigentliche Personalisierungsprozess erfolgt anhand des dynamischen Abgleichs des Nutzerprofiles mit einzelnen möglichen Teilbereichen (Portlets) einer Webseite auf Basis der modellierten Contentkategorien. Laut Schackmann [Scha03] kommen dabei folgende Verfahren zum Einsatz:

- *Elektronische Kataloge* stellen auf Basis der vorgegebenen Suchkriterien Auswahllisten zusammen, aus denen der Benutzer aktiv für ihn geeignete Webseiten aufsucht.

- *Konfigurationsmöglichkeiten* dienen der Voreinstellung der Benutzerpräferenzen und erlauben eine entsprechende automatisierte Anpassung der Webseite bzw. des Informationsangebots.
- *Kontextabhängige Deduktion* bietet den Fachbereichsmitarbeitern die Möglichkeit, Regeln zu definieren, um dem aktiven Benutzer auf Basis seiner Profile entsprechend klassifizierten Inhalt anzubieten.
- *Collaborative Filtering* bestimmt Kunden mit ähnlichen Interessen, entweder durch explizite Nachfrage bei Benutzern oder durch automatischen Vergleich der Bewegungspfade bzw. durch Ermittlung ähnlicher Profile um auf dieser Basis Vorhersagen zu erstellen.
- *Data Mining* (zum Beispiel *Neuronale Netze*) analysiert *User-Log-Files* (meist offline), *User Click-Stream* (meist online) sowie demographische Daten um Verhaltensmuster herauszufinden und Vorhersagen über den relevanten Inhalt zu erstellen.

Wesentliche Unterschiede zwischen den genannten Verfahren liegen in der Fähigkeit zur *Offline-* bzw. *Online-*Verarbeitung, dem Eindringen in die Privatsphäre der Kunden und in der unterschiedlichen Genauigkeit der Personalisierungsergebnisse [Scha03].

3.3.2 Recommender Systeme

Mit dem Begriff *Empfehlungssysteme* bzw. *Recommender Systeme* wird die Software bezeichnet, die vorrangig aus dem Nutzungs- und Kaufverhalten von Besuchern/Kunden einer Webseite geeignete Empfehlungen erzeugt [GaSc02]. Empfehlungssysteme lassen sich laut Literatur [Bohn03] in individualisierte (bzw. personalisierte) und nicht-individualisierte (bzw. nicht-personalisierte) Systeme differenzieren. Nicht-personalisierte Empfehlungssysteme generieren gleiche Empfehlungen für alle Benutzer einer Webseite, in Form von Informationen über Wissenswertes (zum Beispiel TopTen-Listen bekanntester Schriftsteller oder Musiker). Die personalisierten Empfehlungssysteme sind an einzelne Benutzer persönlich gerichtet. Dabei kommen inhaltsbasierte (*content-based*) oder kollaborative (*collaborative*) Filterverfahren zur Anwendung. Im Rahmen einer *automatisierten Empfehlung* können nicht nur Informationen und Produkte empfohlen werden sondern auch andere Personen, die zum Beispiel einer Community angehören. Das heißt es kann auf andere Personen hingewiesen werden, die mit einer Problemstellung weiterhelfen könnten [Koch03]. Im Vergleich dazu empfehlen sich beim *aktiven kollaborativen Filtering* die Benutzer gegenseitig Objekte, welche mit der Push-Kommunikation verbreitet werden.

Recommender Systeme lassen sich in zwei grundlegendste Methoden differenzieren. Die allgemeine Methode wird durch die auf *Server laufenden Anwendungen* verwirklicht, während die spezielle Methode die auf *Clients laufenden Anwendungen* voraussetzt. Die spezielle Methode ist im Vergleich zu der Allgemeinen in der Praxis noch selten vorzufinden. Auch das Semantic Web ist ein noch wenig verbreitetes Paradigma der Informationstechnologie. Dennoch trägt die Entwicklung des Semantic Web zur Verbesserung des aktuellen Webs bei, in dem es die Optimierung der Web-Dienste ermöglicht.

Demselben Zweck dienen auch die auf Clients laufenden Anwendungen. Zudem nimmt die Gewichtung hinsichtlich ihrer Relevanz auch mit Semantic Web zu. Im Zusammenhang mit dem Semantic Web sind auf jeden Fall auch weiterführende Optimierungsmöglichkeiten zu berücksichtigen, so dass im Folgenden auf die spezielle Methode näher eingegangen wird.

Spezialfall: Client-Anwendung

Die personalisierten Produktempfehlungen lassen sich anhand eines Vergleichs zwischen persönlichen Präferenzen eines Benutzers und ähnlichen Präferenzen anderer Benutzer erstellen. Diese existierende Informationsstruktur der personalisierten Produktempfehlung lässt sich laut Ziegler [Zieg05] durch eine lokal laufende Anwendung ausschöpfen. Die personalisierte Empfehlung könnte lokal auf dem *Personal Computer* (PC) des Benutzers laufen, um Daten aus dem gesamten *Semantic Web* zu sammeln. Das Empfehlungssystem würde laut Ziegler [Zieg05] lediglich folgende Merkmale aufweisen:

- *Zentralisierte Schätzung*: Alle Berechnungen und Schätzungen werden auf einem einzigen Ort durchgeführt, nämlich lokal auf dem PC des betreffenden Benutzers. Dadurch entstehen Beschränkungen in Bezug auf die Leistung des einzelnen PCs. Andererseits wiederum benötigt der einzelne PC des Benutzers nicht dermaßen viel Kapazität als ein high-speed Server, der mehrere Tausende Benutzerprofile verwaltet und mehr als Tausend hoch n Schätzungen ausführt.
- *Dezentralisierte Datenlagerung*: Obwohl die Berechnungen lokal ablaufen, ist dies mit Daten und Informationen über Präferenzen nicht der Fall. Die relevanten Daten werden im gesamten Netz verteilt. Das Netz wiederum enthält nur Teilansichten seiner Umwelt.

Bild 8 zeigt die Methoden der individualisierten Empfehlungssysteme, welche sich in *Content-based Filtering* und *Collaborative Filtering* unterteilen lassen. Die Mischform der beiden Methoden nennt man *Erweiterte Empfehlungssysteme* bzw. *Hybride Systeme*. In diesem Zusammenhang werden die genannten Methoden beschrieben.

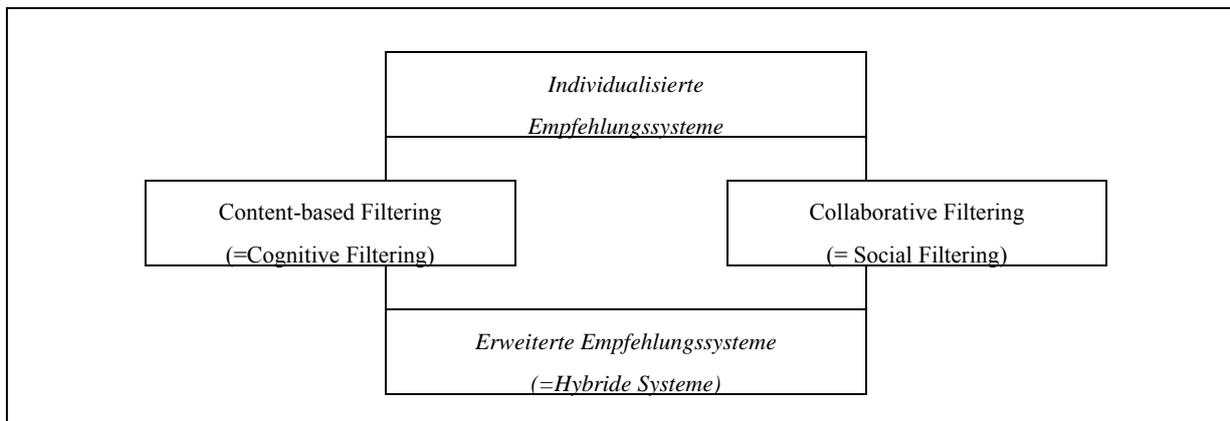


Bild 8 Methoden der individualisierten Empfehlungssysteme

Content-based Filtering berechnet Ähnlichkeiten zwischen den Präferenzen der Benutzer und der Gesamtheit aller Produkte. *Content-based Filtering* funktioniert nur im Zusammenhang mit Domains wenn die Herkunft des Attributs sichtbar ist und die relevanten Informationen zu den Attributen zur Verfügung stehen [Zieg05].

Collaborative Filtering berechnet Ähnlichkeiten zwischen Benutzern bzw. zwischen den Bewertungsprofilen der in Beziehung gesetzten Benutzer. Solange man von der begründeten Annahme ausgeht, dass die Benutzer mit ähnlichen Profilen auch ähnlichen Geschmack in Bezug auf die bestimmte Produktpalette aufweisen, wird ihnen auch eine Beraterfunktion zugewiesen. Daher werden die Profile dieser Benutzer für die Erzeugung von Empfehlungen in Bezug auf die relevantesten Produkte herangezogen [Zieg05]. Ein Praxisbeispiel lässt sich an *amazon.com* illustrieren. Zum Beispiel werden Empfehlungen in der folgenden Art dargestellt: „*Andere Leser, die das selbe Buch gekauft haben, waren auch an folgenden Büchern interessiert ...*“

Collaborative Filtering (CF) verwendet weniger die inhaltliche Darstellungsweise sondern vielmehr die benutzerbasierte Darstellungsform, indem die Benutzer den entsprechenden Produkten Werte zuordnen. Somit führt der Benutzer auf freiwilliger Basis die Bewertungen durch. CF bedient sich daher der Bewertungen, die direkt vom Benutzer abgegeben werden. Definitionsgemäß lässt sich argumentieren, dass der CF Algorithmus auf einer Menge von Benutzern $A = \{a_1, a_2, \dots, a_n\}$, einer Menge von Produkten $B = \{b_1, b_2, \dots, b_m\}$ und partiellen Bewertungsfunktionen $r_i : B \rightarrow [-1, +1, ?]$ operiert. Diese Funktion wird für jeden Benutzer a_i aus der Gesamtmenge A herangezogen. Negative Ergebniswerte der Bewertungsfunktion $r_i(b_j)$ bedeuten Abneigung gegenüber dem Produkt b_j , während positive Werte die Vorliebe für das Produkt b_j ausdrücken. Gilt für die Bewertungsfunktion $r_i(b_j) = ?$ wird mit dem Fragezeichen initialisiert, dass der Benutzer a_i das Produkt b_j noch nicht bewertet hat. Collaborative Filtering Systeme kennzeichnen sich insbesondere durch hohe Outputqualität und durch minimale Informationsanforderung. Laut Literatur [Zieg05] sind sie auch die bedeutendsten Vertreter der Empfehlungssysteme.

Ansätze des *Bayesian Classifiers*, *Horting* und *Association Rule-based Techniques* wurden zur Realisierung des *Collaborative Filtering* eingesetzt. Die weiteste Verbreitung jedoch gelang den Ansätzen des *User-based* und *Item-based Collaborative Filtering* [Zieg05].

Das *Collaborative Filtering* weist gegenüber dem *Content-based Filtering* einige Vorteile auf. Die Vorteile des *Collaborative Filtering* sind insbesondere die Unabhängigkeit von Produkteigenschaften sowie die Unabhängigkeit der Prognosegenerierung von nicht-ähnlichen oder neuen Produkten. Ein häufiges Hindernis des *Collaborative Filtering* stellt insbesondere das Cold-start Problem dar, das immer dann auftritt, wenn die Benutzerprofile leer sind und daher die erforderlichen Daten nicht vorhanden sind. Denn es können nur jene Produkte empfohlen werden, die bereits bewertet wurden [soko05].

Erweiterte Empfehlungssysteme tendieren dazu *Collaborative* und *Content-based Filtering* zu vereinigen um deren Nachteile zumindest partiell zu beseitigen, indem entweder am Ansatz selbst oder an den Synergieeffekten angeknüpft wird. Aufgrund dieser Mischform werden diese Systeme auch hybride Systeme genannt [Zieg05].

Im Bereich der *hybriden Systeme* wurden auch zahlreiche Methoden entwickelt. Die weitverbreitetste Methode ist die *Collaboration-via-content*, anhand derer die Ähnlichkeiten zwischen den Benutzern anhand der Zufriedenheit der Benutzer eruiert werden. Dabei werden Profile erzeugt, welche die Zufriedenheit des Benutzers abbilden und anhand dieser Profile werden ähnliche Benutzerprofile festgestellt [Zieg05].

Collaborative Filtering

Die Durchführung von Personalisierung anhand des *Collaborative Filtering* basiert auf der Ähnlichkeitsschätzung zwischen den Benutzern [soko05]. In der Literatur wird zwischen *Active* und *Passive (=Automated) Collaborative Filtering* unterschieden, wobei beim *Active Collaborative Filtering* der Benutzer *aktiv* Objekte weiterempfiehlt. Somit tauschen sich die Benutzer gegenseitig Objekte aus. In weiterer Folge wird jedoch ausschließlich auf *Automatic Collaborative Filtering* eingegangen, also auf jene Methoden welche sich durch automatische Datenerhebung und durch automatische Ableitung der fehlenden Präferenzen charakterisieren lassen. Das automatische Benutzermatching kann entweder anhand des *Model-based Collaborative Filtering* oder anhand des *Memory-based Collaborative Filtering* durchgeführt werden [soko05].

Das *Model-based CF* basiert auf einem zuvor geschätzten Modell, während das *Memory-based CF* aufgrund eines direkten Vergleichs der Benutzerprofile operiert. Beim *Memory-based CF* werden die Benutzerdaten entweder implizit oder explizit erhoben. Implizite Datenerhebung kann anhand von Transaktionsdaten und Clickstream-Analysen erfolgen, während bei der expliziten Datenerhebung die vom Benutzer diesem Zweck direkt abgegebenen Bewertungen herangezogen werden. Für die Analyse der relevanten Daten werden Proximitätsberechnungen durchgeführt. Das dabei verwendete Proximitätsmaß (*cosine-based similarity, correlation-based similarity, adjusted cosine similarity*) wird für die Erstellung von Prognosen bzw. Empfehlungen herangezogen [soko05].

Mit dem *Model-based Collaborative Filtering* ist eine genauere Vorhersage möglich, denn sobald ein Modell erstellt ist, kann es zur gezielteren Profilerhebung eingesetzt werden. Dabei kommen modellbasierte Algorithmen wie Cluster-Modelle, Bayessche Netze oder Neuronale Netze zur Anwendung [soko05].

In der Literatur wird für *Model-based Collaborative Filtering* häufig der Begriff *Item-based Collaborative Filtering* verwendet, da der Synonymname auf das Objekt fokussiert. Statt *Memory-based Collaborative Filtering* wird synonym der Begriff *User-based Collaborative Filtering* eingesetzt, da der Name auf den Benutzer fokussiert und die Methode durch die Synonymbezeichnung genauso gut beschrieben wird. In weiterer Folge wird aufgrund der charakteristischen Bezeichnung mit den Synonymen gearbeitet.

Einige der ersten Empfehlungssysteme, welche die *User-based Collaborative Filtering Methode* anwendeten, wurden in Ringo und GroupLens-Projekten realisiert. Bei dieser Methode wird Bewertungsfunktion r_i jedes Benutzers a_i als Vektor dargestellt sowie die Ähnlichkeiten zwischen allen Paaren $(a_i, a_j) \in A \times A$ berechnet. Diesem Zweck werden allgemeine statistische Korrelationskoeffizienten wie die *Pearson Correlation* und die *Cosine Similarity* angewendet. *Cosine Similarity* misst die Ähnlichkeit zwischen zwei Vektoren anhand des Kosinuswertes ihrer Winkel. Die *Pearson Correlation*, die vom linearen Regressionsmodell abgeleitet wird, ist zwar der *Cosine Similarity* ähnlich, jedoch misst sie in wieweit eine lineare Beziehung zwischen zwei Variablen vorhanden ist. Dabei wird mit Durchschnittswerten der Vektoren gerechnet. Mit diesen Methoden werden die ähnlichsten Nachbarn berechnet. Daraufhin werden Empfehlungen für alle Produkte berechnet, die von den Nachbarn eines bestimmten Benutzers bewertet wurden, jedoch nicht von diesem Benutzer selbst [Zieg05]. Der Benutzer für den Empfehlungen generiert werden, wird in der Literatur häufig *aktiver Benutzer* genannt, da eine direkte Empfehlung auch nur dann abgegeben werden kann, wenn der Benutzer auch aktiv ist und sich auf der Webseite des Unternehmens befindet.

Item-based Collaborative Filtering ist in den letzten fünf Jahren aufgrund der Vorzüge der günstigen Berechnungsmethoden sowie aufgrund der Entkopplung des Modellberechnungsprozesses von der aktuellen Vorhersagegenerierung in Schwung gekommen. Im Vergleich zu *User-based Collaborative Filtering* werden hierbei die Ähnlichkeitswerte nicht zwischen den Benutzern a_i , sondern zwischen Objekten bzw. Produkten berechnet, daher ist die Methode des *Item-based CF* auch objekt- bzw. produkt-fokussiert. Wenn nun zwei Objekte im Vergleich zueinander ähnlich sind, das heißt wenn sie einen hohen Ähnlichkeitswert aufweisen, werden ihnen von zwei Benutzern identische oder ähnliche Werte zugewiesen. Bei positiver Korrelation werden die Objekte von den Benutzern bevorzugt [Zieg05].

Empfehlungssysteme gewinnen immer weiter an Bedeutung aufgrund ihrer Einsatzflexibilität und aufgrund ihrer bedeutenden Funktion als Komplexitätsminimierer für den aktiven Benutzer. In der informationstechnologischen Literatur wird bereits der Weg in die Richtung von sozialeren und network-orientierten Empfehlungssystemen eingeschlagen. Als Grundlage dessen dienen vertrauensbasierte und dezentralisierte Empfehlungssysteme [Zieg05].

Praxisbeispiel: *MusicStrands*

Ein Praxisbeispiel für ein Unternehmen, das sich mit der Entwicklung der Personalisierungssoftware beschäftigt, ist *MusicStrands* (<http://indy.musicstrands.com/corp/products/recommender.html>). *MusicStrands Recommender™* Produkte fördern *one-to-one* Personalisierung für Unternehmen, die sich in Musikbereich spezialisieren, indem sie die Präferenzen der einzelnen Benutzer für Musik abbilden [MuSt05].

Recommender Engine™ ist ein Software System, das fortgeschrittene *Artificial Intelligence Verfahren* verwendet um die Empfehlung von *Songs* und *Playlists* für Benutzer oder Gruppen von Benutzern zu realisieren. Die *Recommender Engine™* ist dermaßen konstruiert, dass sie in Anwendungen direkt integriert werden kann oder von der existierenden Anwendung aufgerufen werden kann. Die Software funktioniert nur mithilfe notwendiger Datenstruktur, die durch Daten über Songs und entsprechenden Daten über Präferenzen der einzelnen Benutzer abgebildet wird. Die Anwendungen des einzelnen Unternehmens kommunizieren mit dem Recommender System via URLs. Das Unternehmen hat dabei die vollständige Kontrolle über die Datensätze, über die Feinabstimmung des Systems sowie über die Ausgabe des Outputs [MuSt05].

Recommender IT™ stellt für den Programmierer ein Werkzeugsatz dar, welches das Recommender System sinnvoll ergänzt. Anhand des Werkzeugsatzes können Teilbereiche des Recommender Systems auch fernbedient werden, nämlich auf Basis von Web-Dienstleistungen, die mit der Syntax der aktuellen Web-Standards übereinstimmen. Wenn die bevorzugten Software-Werkzeuge der Programmierer von den aktuellen Web-Standards differenzieren, kann die Standard-Syntax der *Web Services Description Language* (WSDL) herangezogen werden um die einfließenden Anfragen und Reaktionen in die neuen Anwendungen und Webseiten einzubeziehen. [MuSt05]

Die vollständige WSDL 1.2 Syntax ist im Anhang dieser Arbeit dargestellt. In diesem Zusammenhang wird auf die Webseite des World Wide Web – Consortium, <http://www.w3.org> verwiesen, wo die zahlreichen Codes und Standards nachgelesen werden können.

Anwendungsbereich

„[...] die vielfältigen Einsatzmöglichkeiten von Recommender Systemen lassen erkennen, dass mit dieser Technologie natürlich Web-Controlling Aufgaben bearbeitet werden können.“ [GaSc02]

Das „Web-Controlling“ steht synonym für die „Erfolgskontrolle im Web“. In der Literatur finden sich unterschiedliche Definitionen des Begriffs, welche im Grunde dasselbe aussagen. Eine der kompakten Definitionen ist die von Cognos GmbH, einem im Bereich *Business Intelligence* und *Performance Management* tätigen Unternehmen: *„Web-Controlling steht für methodische Effizienzkontrolle und Effizienzsteuerung von Web-Auftritten durch systematisches Auswerten von Kunden- und Nutzungsdaten.“ [Cogn00]* In vielfältiger Weise kann bei der Erfolgskontrolle vorgegangen werden. Dabei werden immer aus informationsarmen Daten aufschlussreichere Informationen gewonnen um die Erfolgskontrolle im Web zu gewährleisten [Spil01; HKMW01]. Die dabei zur Anwendung kommenden Methoden erstrecken sich daher von einfachen *Data Mining Methoden* bis zu *Semantic Web Mining*. Das Recommender System stellt eine alternative und auch wirkungsvolle Methode dar und findet daher zunehmend Anwendung in diesem Bereich.

Nachfolgend wird auf die vielfältigen Einsatzmöglichkeiten von Recommender Systemen eingegangen. Dabei werden laut Literatur [GaSc02] entsprechende Einsatzmöglichkeiten sowohl auf der Seite der Benutzer als auch auf der Seite der Anbieter identifiziert.

Gaul et al. [GaSc02] identifizierten auf der Seite des Benutzers (bzw. Nachfrager-Seite) bestimmte Einsatzmöglichkeiten der Recommender Systeme. Dabei unterstützen die Recommender Systeme die:

- Tendenz, die Webseiten-Besucher zu Käufern zu machen
- Verbesserung von Cross- und Up-Selling Möglichkeiten
- Verfestigung der Kunden-Loyalität

Auf der Anbieter-Seite (bzw. Webseiten-Betreiber) lassen sich die Einsatzmöglichkeiten insbesondere in der nutzenstiftenden Funktion der Recommender Systeme erkennen. Laut Gaul et al. [GaSc02] helfen Recommender Systeme bei der:

- Klassifikation von Konsumentenbedürfnissen
- Entdeckung von Cross- und Up-Selling Möglichkeiten
- Überprüfung von Nachfrage-Trends

Durch gezielte Kennzahlengenerierung können Recommender Systeme dazu beitragen, den Web-Auftritt des Unternehmens zu optimieren [GaSc02]. Dabei werden Daten über Kunden mit Web Mining Methoden eruiert. Diese Daten werden von Empfehlungssystemen zu brauchbaren Informationen verarbeitet um darauf hin dem Kunden personalisierte Inhalte zu liefern. Der Kunde erhält somit Informationen über Produkte, die ihn zu großer Wahrscheinlichkeit ansprechen und daher zum Kauf anregen. Der Kunde wird auf diese Weise mit Informationen versorgt, die sein Interesse wecken und ihn zum Kauf bewegen können.

4 ANALYSE

In diesem Kapitel werden die Einflussfaktoren hinsichtlich ihrer Relevanz *im Allgemeinen* für die Praxis und *im Speziellen* für den Anwender und für das Unternehmen untersucht. In der Praxis kommt das Web generell aus unterschiedlicher Motivation heraus zur Anwendung. Hauptsächlich lassen sich die Aktivitäten in zwei Gruppen zusammenfassen: jene des Absatzes bzw. Erwerbs von Produkten und sonstige Aktivitäten. Ersteres gehört zum Bereich der digitalen Wertschöpfung (Definition des Begriffs in Kapitel 3.1) während sonstige Aktivitäten zwecks Unterhaltung, Bildung oder Information entstehen. Somit stellt die erste Gruppe die der wirtschaftlichen Aktivitäten dar, welche die Käufer-Verkäufer Situation voraussetzt. In der zweiten Gruppe sind Anwender des Webs, für welche lediglich eine unbestimmte Aktivität im Web vorausgesetzt wird. Der Anwender kann ein potentieller Kunde sein, er kann aber auch nur ein einfacher Besucher der Webseite sein. In der Praxis werden die Aktivitäten beider Gruppen berücksichtigt, daher wird die Behandlung der Einflussfaktoren in der Praxis zum allgemeinen Fall erklärt. Die Spezialfälle *Anwender-Sicht* und *Unternehmer-Sicht* können dagegen eine Vielzahl von Aktivitäten ausschließen bzw. sich auf spezielle Aktivitäten im Web beschränken. Die Analyse ist in vier wesentliche Subkapitel gegliedert, welche im Bild 9 illustriert sind.

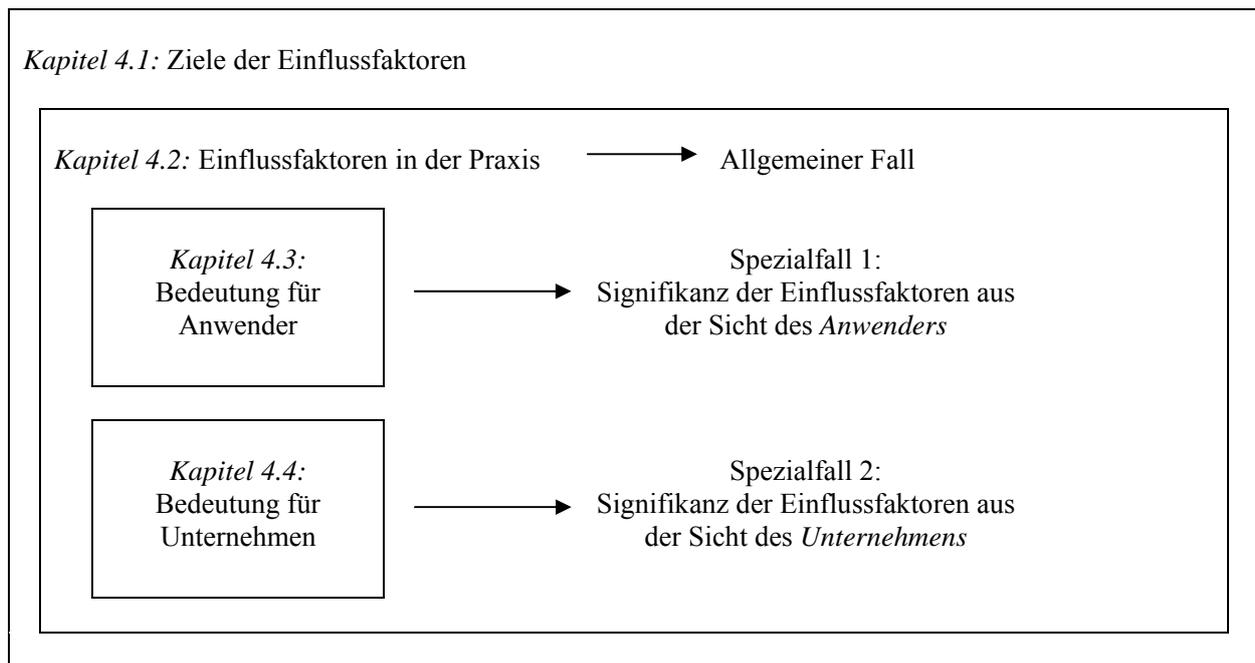


Bild 9 Kapitel-Übersicht

Kapitel 4.1 stellt eine Differenzierung der Einflussfaktoren auf, indem jeweils das wesentliche Merkmal hervorgehoben wird. Auf dieser Basis werden die Ziele der genannten Einflussfaktoren erklärt und graphisch dargestellt.

Kapitel 4.2 dient der praxisorientierten Darstellung der Einflussfaktoren. Kapitel 4.2.1 behandelt die wesentlichen technologischen und unternehmerischen Aspekte des Mobilfunknetzes. Mit dieser kurzen Einführung in den Bereich des Mobilfunknetzes wird gezeigt wie das mobile Telefonieren technisch möglich ist, das heißt wie das Handynetz funktioniert, welche Antennentypen es gibt und wie ein Mobilfunknetz aufgebaut ist. Anschließend werden zwei Unternehmen des Mobilfunknetzes in Österreich präsentiert, das heißt der Aufgabenbereich und die Unternehmensphilosophie der Mobilkom Austria und der One GmbH werden zum Thema. Kapitel 4.2.1 gibt somit einen praxisnahen Überblick über das Mobilfunknetz sowohl aus der technischen Sicht als auch aus der unternehmerischen Sicht. In Kapitel 4.2.2 wird auf der Grundlage dieser praxisnahen Darstellung des Mobilfunknetzes die Signifikanz der Einflussfaktoren für die Praxis dargestellt. Das Kapitel 4.2.3 stellt die wesentlichen Aspekte der Optimierung von kommerziellen Webseiten dar. Das Ziel besteht darin die einfache Orientierung für den Anwender auf der Webseite zu gewährleisten. Die drei speziellen Methoden der Webseitenoptimierung nach Bensberg zeigen wie eine kommerzielle Webseite eines online-tätigen Unternehmens optimiert werden kann.

Anschließend wird in Kapitel 4.3 die Bedeutung der Einflussfaktoren für die Anwender anhand einiger Beispiele erläutert. Aufbauend auf der theoretischen Grundlage der Einflussfaktoren werden in Kapitel 4.4 die informationstechnologischen Instrumente, *Semantic Web Mining* und *Recommender Systeme* hinsichtlich ihrer wirtschaftlichen Bedeutung für Unternehmen untersucht.

4.1 Ziele der Einflussfaktoren

Im Folgenden werden die Einflussfaktoren entsprechend ihrer Fokussierung anhand von Ebenen differenziert. Die Instrumente, die dabei zum Einsatz kommen, werden entsprechend zugeordnet und von jedem Einflussfaktor wird das primäre Ziel abgeleitet.

Die drei Einflussfaktoren *Technologie für das Web*, *Information aus dem Web* und *Personalisierung im Web* fokussieren auf je einen speziellen Bereich, wie der Name jeweils andeutet: der erste Einflussfaktor fokussiert auf **T**echnologie, der Zweite auf **I**nformation und der Dritte auf **P**ersonalisierung. Die Einflussfaktoren werden in weiterer Folge auch mit der Abkürzung „TIP-Faktoren“ verwendet. Dementsprechend werden auch technologie-, informations- und personalisierungsorientierte Instrumente der Informationstechnologie eingesetzt. Auf der Technologie-Ebene befindet sich der Ansatz des *Semantic Web* als bedeutendstes Instrumentarium der Informationstechnologie. Auf der Informations-Ebene ist das Konzept des *Semantic Web Mining* in dieser Arbeit gewähltes Instrument der Informationstechnologie. Auf der Personalisierungs-Ebene werden *Customer Relationship Management* und insbesondere die Personalisierungssoftware, *Recommender Systeme* als ökonomisch effektive Instrumente herangezogen. Bild 10 stellt die genannten TIP-Faktoren und die informationstechnologischen Instrumente graphisch dar.

Auf Basis dieser drei Ebenen, die entsprechend ihrer Fokussierung benannt sind, wird nun auf die Ziele der Einflussfaktoren eingegangen.

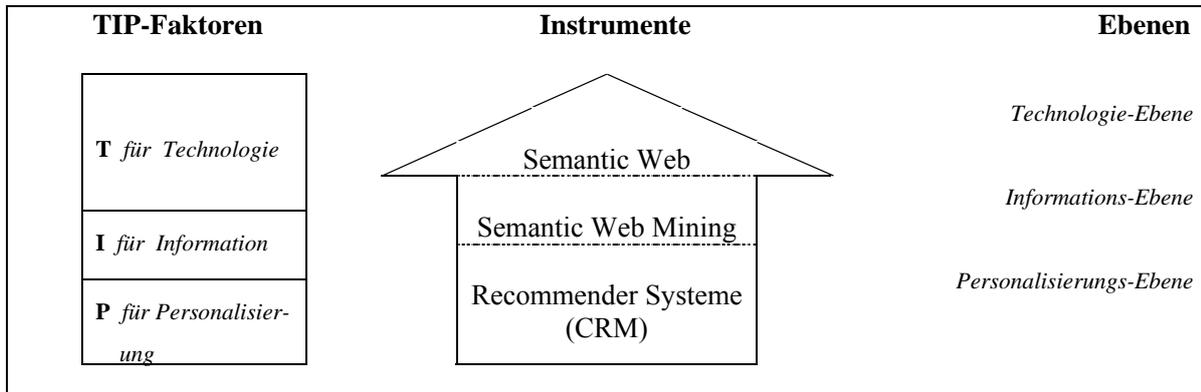


Bild 10 Instrumente der Informationstechnologie auf drei Ebenen

Aus den drei Einflussfaktoren lassen sich drei Ziele ableiten. Der erste Einflussfaktor *Technologie für das Web* ist durch die Förderung von Forschung und Entwicklung realisierbar und soll durch die Innovation des Semantic Web, dem Ziel ein maschinenfreundlicheres Web zu gestalten, gerecht werden. Der zweite Einflussfaktor *Information aus dem Web* dient dem Ziel ein benutzerfreundlicheres Web zu gestalten, indem der Fokus auf die Information gelegt wird. Der dritte Einflussfaktor *Personalisierung im Web* ist zum Zweck der Unternehmensphilosophie, *Customer Relationship Management*, bedeutendes Instrument zur Erreichung des dritten Zieles, nämlich ein kundenfreundlicheres Web zu gestalten. Bild 11 stellt die zu den Einflussfaktoren gehörenden Ziele graphisch dar. Zudem bauen die Ziele aufeinander auf: ein maschinenfreundliches Web unterstützt die Benutzerfreundlichkeit des Webs; ein maschinen- und benutzerfreundliches Web trägt wiederum zu einem kundenfreundlicheren Web bei.

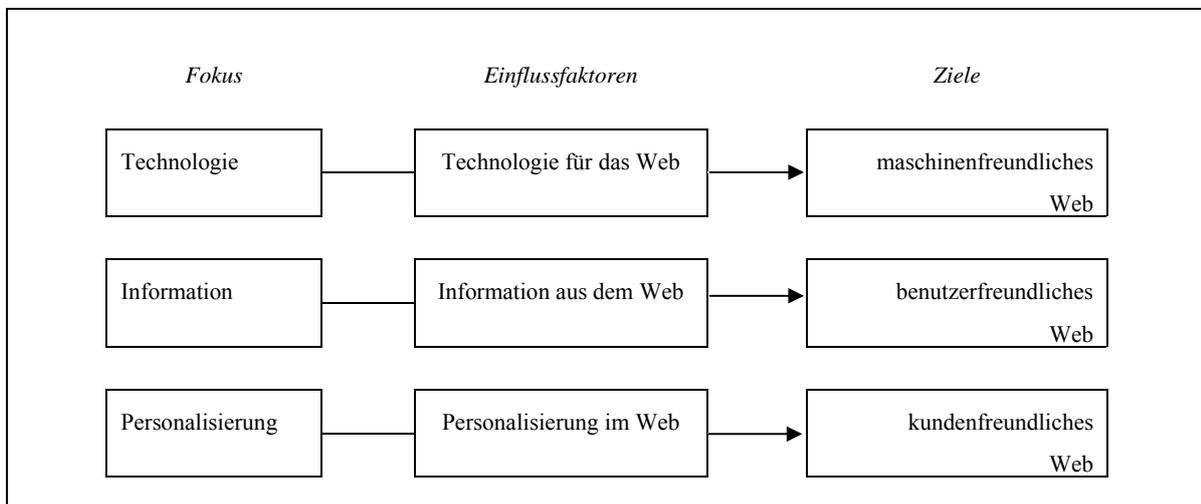


Bild 11 Ziele der Einflussfaktoren

Aus den genannten Zielen lässt sich auch die wirtschaftliche Bedeutung der Einflussfaktoren ableiten. In Kapitel 4.4 wird anhand der präsentierten Einflussfaktoren die wirtschaftliche Bedeutung insbesondere der informationstechnologischen Instrumente *Semantic Web Mining* und *Recommender Systeme* erörtert.

4.2 Signifikanz der Einflussfaktoren für die Praxis

„Der Synergie-Grundsatz, dass das Ganze mehr wert sei als die Summe seiner Teile, verdient bei der Internet-Strategie ganz besondere Aufmerksamkeit.“ [Steu98]

Anhand der repräsentierten Einflussfaktoren und dargestellten technologischen Instrumenten lässt sich abstrakt eine Pyramide dieser Konzepte darstellen und begründen. Die erforderliche Begründung lässt sich aus zwei Perspektiven erläutern, einerseits aus dem *Bottom-Up* Blickwinkel und andererseits aus der *Top-Down* Perspektive.

Die *Technologie für das Web* stellt die Grundlage für weiteren Einsatz von informationstechnologischen Instrumenten dar. Auf der Basis des semantischen Webs können *Informationen aus dem Web* effizient gesammelt werden, indem das Semantic Web Mining eingesetzt wird. Die dabei gespeicherten Informationen lassen sich von Recommender Systemen weiterverarbeiten um schließlich die erforderliche *Personalisierung im Web* zu realisieren, indem für den einzelnen Kunden personalisierte Empfehlungen erstellt werden. Diese Begründung beschreibt die *Bottom-Up Perspektive*.

Bild 12 stellt die Pyramide der informationstechnologischen Instrumente graphisch dar.

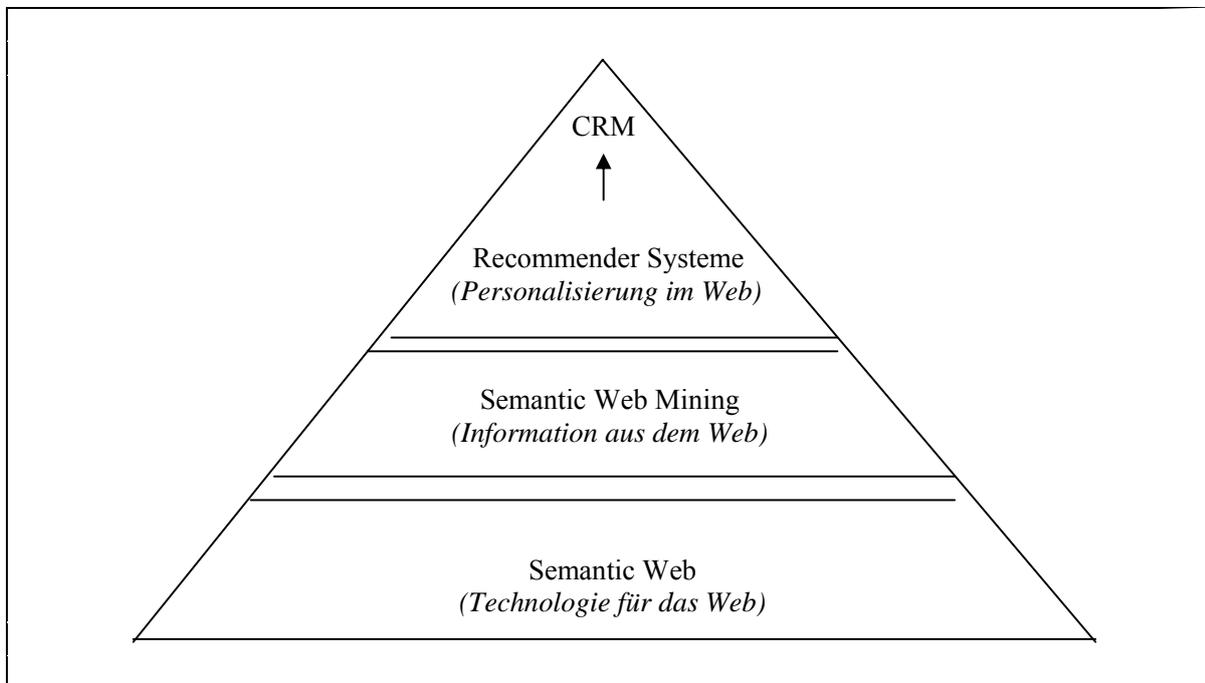


Bild 12 Pyramide der informationstechnologischen Instrumente

Aus der *Top-Down Perspektive* lässt sich die Pyramide der Konzepte folgendermaßen begründen. Customer Relationship Management (CRM) im Web setzt den Einsatz von Web Mining Methoden voraus um Informationen sammeln und aufbereiten zu können. Auf Basis dieser Information lässt sich die erfolgreiche Strategie der Kundenbeziehung realisieren. Die notwendigen Informationen können zum Beispiel von Recommender Systemen verarbeitet werden um personalisierte Empfehlungen zu generieren. Semantic Web Mining setzt wiederum das semantische Netz voraus. Die Technologie des Semantic Web ist daher die grundlegendste Voraussetzung für weitere Einflussfaktoren. Auf der stabilen Basis des semantischen Webs, welche durch entsprechende Ontologieumgebungen realisiert ist, lassen sich in der Praxis weitere informationstechnologische Instrumente einsetzen.

Die Realisierung der drei vorgestellten Einflussfaktoren erlaubt die Optimierung der Dienste im Web. Dabei kommen die im Zusammenhang mit den Einflussfaktoren vorgestellten Instrumente der Informationstechnologie zum Einsatz. Die Optimierung der Web-Dienste ist mit der Realisierung des *Semantic Web* und mit dem Einsatz der Instrumente *Semantic Web Mining* und *Recommender Systeme* erzielbar. In wie weit die Web-Dienste dabei optimiert werden und wie sich dies begründen lässt, zeigt das Kapitel 4.2.2 für den *allgemeinen Fall* anhand diverser Beispiele und das Kapitel 4.2.3 für den *speziellen Fall* anhand der Methoden zur Webseiten-Optimierung. Im Zusammenhang mit dem speziellen Fall werden die Methoden zur Optimierung der Webseiten nach Bensberg präsentiert. Einführend wird nun in Kapitel 4.2.1 das Mobilfunknetz zuerst aus der technologischen und danach aus der unternehmerischen Sicht vorgestellt.

4.2.1 Das Mobilfunknetz

„Elektromagnetische Felder sind seit Entstehung der Erde ein Bestandteil unserer Umwelt. Das Erdmagnetfeld, sichtbar in der Ausrichtung einer Kompassnadel, oder das elektromagnetische Feld der Sonne bestimmen seit Urzeiten die Entwicklung des Lebens auf der Erde.“ [MoAu05a]

Im 21. Jahrhundert sind elektromagnetische Wellen aus dem Leben des Menschen nicht mehr wegzudenken. Ein Toaster, Kühlschrank, Rasierapparat oder ein Handy, also jedes technische Gerät erzeugt beim Verbrauch von Strom elektromagnetische Felder. Auch für die Informationsübertragung bei Fernsehen, Radio oder Mobilfunk sind elektromagnetische Felder ein physikalisch unvermeidbares Nebenprodukt. Zudem sind sie das Transportmedium für alle drahtlosen Informationen und Signale [MoAu05a]. Sie stellen die *Verbindung* zum Beispiel zwischen zwei Mobiltelefonen dar. Etwas abstrakter formuliert, stellen sie die Verbindung zwischen Menschen, Ländern und Kontinenten dar.

Wie ein Handynetz funktioniert

Ist das Handy eingeschaltet, steht es in periodischem Kontakt mit den Mobilfunkanlagen. Zur Gewährleistung des Verbindungsaufbaus werden vom Handy und von Mobilfunkanlagen Funkwellen empfangen und gesendet. Telefonieren ist nur möglich, wenn dieser Kontakt besteht. Wenn eine Mobilfunkanlage in der Nähe ist oder wenn das Handy in der Nähe einer Mobilfunkanlage ist, kann der Handy-Besitzer telefonieren. Das Prinzip der Kundenorientierung erfordert eine Anpassung an die Bedürfnisse der Kunden. Der Kunde wird erst dann von der Dienstleistung Gebrauch machen, wenn das Telefonieren an seinem Standort möglich ist. Daher ist die Aufgabe der Mobilkom Austria „*ein flächendeckendes Mobilfunknetz in ganz Österreich sicherzustellen*“ [MoAu05a]. Die Realisierung eines in ganz Österreich flächendeckenden Netzes setzt die gleichmäßige Verteilung von Mobilfunkanlagen voraus.

Aufbau eines Mobilfunknetzes

Base Transceiver Station (BTS) ist die Basisstation und wird im Folgenden auch Mobilfunkanlage genannt. BTS beinhaltet Antennen zur Versorgung einer Funkzelle. Die Sende- und Empfangseinheiten stehen permanent mit jedem eingeschalteten Handy in Verbindung [MoAu05a; mobi05].

Base Station Controller (BSC) dient der Organisation und der Verwaltung von zehn bis hundert solcher Mobilfunkanlagen [MoAu05a].

Mobile Switching Center (MSC) enthält eine Datenbank mit allen Rufnummern und ermöglicht den Kontakt zwischen mehreren Teilnehmern innerhalb verschiedener Mobilfunknetze (Festnetz, One, etc.) [MoAu05a].

Antennentypen

Drei Arten von Mobilfunkanlagen lassen sich anhand der unterschiedlichen Reichweiten identifizieren: die *Picozelle*, die *normale Mikrozelle* und die *Makrozelle*.

Die Picozelle hat die Reichweite von rund 50 Metern und wird daher im dicht verbauten Gebiet der Stadt eingesetzt. Die normale Mikrozelle hat durchschnittlich eine Reichweite von 300 Metern. Die Makrozelle hat eine Reichweite von rund zwei bis drei Kilometern und wird vor allem in ländlichen Regionen eingesetzt [MoAu05a].

In der Regel ist eine Überlappung der einzelnen Funkzellen zu gewährleisten um Funklöcher zu vermeiden. Ist diese Überlappung nicht gewährleistet, können Gespräche abbrechen. Wenn die Funkanlagen nicht die notwendige Reichweite haben oder wenn sie nicht nahe genug aneinander gelegen sind, entstehen Funklöcher in dem Bereich zwischen den Mobilfunkanlagen. Fällt nun ein fahrendes Auto in ein Funkloch, wird das Gespräch des Fahrers abbrechen. Die Funklöcher können durch entsprechend dichte Verteilung der Mobilfunkanlagen bzw. erforderlichen Einsatz von Funkzellen mit entsprechender

Reichweite vermieden werden. Die verschiedenen Standorte der Mobilfunkanlagen bilden im Ganzen ein Mobilfunknetz. Dieses Mobilfunknetz ist umso stabiler, je dichter die Mobilfunkanlagen gelegen und je weniger Funklöcher vorhanden sind. Ist das Mobilfunknetz dicht aufgebaut, ist auch die Sendeleistung zwischen Mobilfunkanlage und Handy beim Gespräch geringer. Im Vergleich dazu kann ein Fernsehsender herangezogen werden. Will der Fernsehsender eine große Region störungsfrei abdecken, erzeugt er eine mehr als *tausendmal größere* Sendeleistung als eine Mobilfunkanlage [MoAu05a].

Mobiles Telefonieren

Unabhängig vom Standort des Handy-Besitzers werden beim Telefonieren schwache Signale an die Antennen der Mobilfunkanlage gesendet, die das Gespräch weiterleitet. Angenommen der Handy-Besitzer telefoniert auf dem Weg zur Arbeit und sitzt gerade in seinem Auto. In diesem Fall verlässt er die Reichweite einer Mobilfunkanlage (BTS) und wird mit seinem Gespräch an die nächste weitergeleitet [MoAu05a].

Die Vermittlungsstelle (MSC) hat eine angeschlossene Datenbank, in der sich unter anderem die Informationen über den momentanen Standort aller aktiven Handys befinden. Dadurch können Gesprächsverbindungen erst aufgebaut werden. Die Daten dieser Datenbank unterliegen dem Datenschutz [MoAu05a]. Bild 13 zeigt wie die Gesprächsverbindung zwischen zwei Personen hergestellt wird und wie die Vermittlungsstelle zwischen dem Anrufer und Angerufenen (das heißt zwischen der Mobilfunkanlage X und Y) agiert.

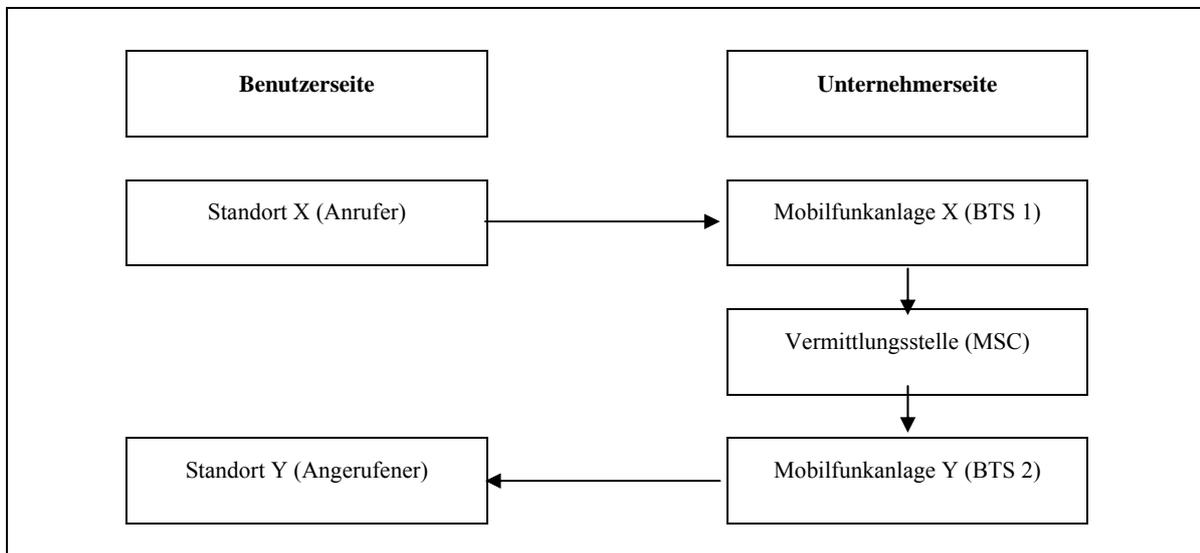


Bild 13 Gesprächsverbindung zwischen zwei Handy-Benutzern

Abgesehen vom Gesprächsaufbau werden auch andere Funktionalitäten, die insbesondere vom Endgerät abhängig sind, über BTS übertragen. In Verbindung mit den gängigen Netzstandards wie *Global System for Mobile Communications (GSM)*, *General Packet Radio Service (GPRS)*, *Universal Mobile Telecommunications System (UMTS)*, *Enhanced Data Rates for Global Evolution (EDGE)* und den

Datenprotokollen wie WAP und i-mode können weitere Funktionalitäten realisiert werden. Unter der Annahme, dass die Endgeräte die entsprechenden Features unterstützen, können Basisfunktionalitäten wie *Short Message Service* (SMS), *Enhanced Message Service* (EMS) oder *Multimedia Messaging Service* (MMS) zur Anwendung kommen [Clem03].

Zu den ortsabhängigen Informationsdiensten gehören *Location Based Services* (LBS), die auf Basis des aktuellen Aufenthaltsortes eines Teilnehmers entsprechende ortsrelevante Informationen liefern. Durch die einfache Zellordnung von GSM-Netzen lässt sich die ungefähre geographische Position des Teilnehmers bestimmen [Clem03].

Standortbezogene Dienste lassen sich zur ortsgebundenen Kundenansprache oder als Suchsysteme sinnvoll einsetzen. Beispiele dafür sind das Aufspüren des nächstgelegenen Bankautomaten, einer nahen Apotheke oder des Programmkinos. Für die Kunden stellen diese Dienstleistungen zusätzlichen Mehrwert dar, für das Unternehmen gleichzeitig zufriedener Kunden. Diese Mehrwertleistungen werden insbesondere von suchenden Kunden in Anspruch genommen, die zum Beispiel nach Adressen suchen oder sich für Prämien und der gleichen interessieren [Clem03].

Unternehmen des Mobilfunknetzes

Ein Telekommunikationsunternehmen verbindet Menschen, in dem er Dienstleistungen und effiziente Produktlösungen anbietet. Gleichzeitig bietet er im Vergleich zu den klassischen Kommunikationsmedien ein Moderneres an, womit nicht nur telefoniert werden kann, sondern auch E-Mail lesen und einkaufen via Handy möglich ist. Für den Endbenutzer heißt das mehr Flexibilität, Varietät und mehr Möglichkeiten der Alltagsgestaltung.

Ein Unternehmen, das aus ökonomischer Motivation heraus mit den Instrumenten der Informationstechnologie fungiert, erfüllt im Web wesentliche wirtschaftliche Funktion. Durch den Einsatz von Semantic Web Mining und Recommender Systemen wird die Grundlage für wirtschaftliche Aktivitäten im Web zudem verbessert.

Das Unternehmen der Telekommunikationsbranche steht zwischen der Web-Technologie und dem Benutzer. Für das Unternehmen bedeutet dies, dass es zwei wichtige Ziele verfolgt. Auf der technologischen Seite (rechte Seite im Bild 14) fördert es die technologische Forschung und Entwicklung (F&E) und auf der anderen Seite (linke Seite im Bild 14) pflegt es die Kundenbeziehung, indem es die Grundsätze der Unternehmensphilosophie des *Customer Relationship Managements* (CRM) anwendet.

Das Potential der Telekommunikationsbranche in der heutigen Gesellschaft liegt im technologischen Fortschritt, sowohl hardware- als auch softwaremäßig. Mit der zusätzlichen Webverbindung, die durch mobile Endgeräte realisiert werden kann, entstehen weitere und insbesondere vielfältigere Möglichkeiten des Informationszeitalters.

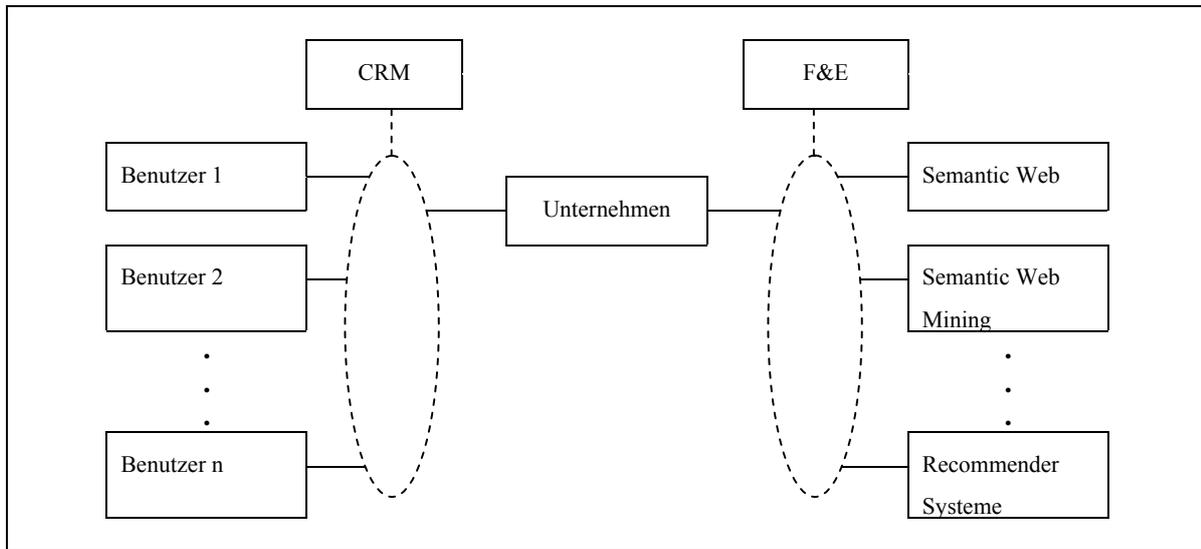


Bild 14 Unternehmen als Schnittstelle zwischen Technologie und Benutzer

Nachfolgend werden zwei der größten Netzproviders in Österreich vorgestellt: *ONE GmbH* und *Mobilkom Austria AG & Co KG*.

ONE GmbH

Eigentümer der ONE GmbH sind die deutsche E.ON AG (<http://www.eon.com/>), die norwegische Telenor (<http://www.telenor.no/>), die französische Orange (<http://www.orange.com>) und die dänische TDC (<http://tdc.dk/>). Seit dem Start am 26. Oktober 1998 hat ONE gemeinsam mit anderen Marktteilnehmern kundenorientierte Produkte auf den Markt gebracht und für innovatives Service gesorgt [OneG05].

Die Unternehmensphilosophie von ONE basiert auf dem Grundgedanken des Customer Relationship Managements, wobei die Zufriedenheit des Kunden und die Beziehung zum Kunden zu den wichtigsten Zielen des Unternehmens gehören. Anhand der Leitgedanken von ONE werden die Ziele wie *einfach*, *zuverlässig*, *herausfordernd*, *kundenorientiert* und *mit dem Herzen dabei* verfolgt. Im Folgenden werden die Leitgedanken von ONE beschrieben.

Der Leitgedanke *einfach* bezieht sich auf die Bedienbarkeit der ONE Produkte und impliziert einfache Lösungswege. Diese Einfachheit wird durch intelligente Anwendungen realisiert, wodurch mit einem einzigen Knopfdruck auf spezifische Dienste zurückgegriffen werden kann [OneG05].

Der Leitgedanke *zuverlässig* impliziert die zuverlässige Netzqualität, wodurch Kunden von ONE sich auf das Netz von ONE verlassen können.

Der Leitgedanke *herausfordernd* bedeutet, dass ONE seine innovativen Erfolge zur Mehrwert-Generierung für den Kunden einsetzt.

Die Leitgedanken *kundenorientiert* und *mit dem Herzen dabei* sein, implizieren die Grundidee der Unternehmensphilosophie, *Customer Relationship Management (CRM)*.

Mobilkom Austria AG & Co KG

„Think global – act local lautet die Strategie von Mobilkom Austria.“ [MoAu05b]

Mobilkom Austria AG & Co KG ist eine Tochter der Telekom Austria. 1992 ging sie aus dem Funktechnischen Dienst (FTD) hervor. Mobilkom Austria zählt zu den größten Mobilfunkbetreibern in Österreich. Mit dem Vertrieb von GSM-, UMTS- und EDGE-Diensten agiert sie unter der Marke A1 (daher auch der Name der Webseite: <http://www.a1.net>) auf dem österreichischen Mobilfunk-Markt.

Zu der mobilen Kommunikation der Mobilkom Austria zählen Sprachtelefonie, alle Arten der mobilen Datenübertragung, mobile Zahlungslösungen und der Bereich *Customer Services*. Das wichtigste Geschäftsfeld der Telekom Austria ist zwar noch immer die Sprachtelefonie. Durch die Innovationen in der Technologie und im Service werden zukünftig auch erweiterte Dienste der mobilen Datenübertragung wie GPRS, EDGE und MMS zu den bedeutenden Geschäftsfeldern der Telekom Austria gehören [MoAu05b].

Mobilkom Austria AG & Co KG beteiligt sich an Mobilfunkunternehmen in den aufstrebenden Märkten Südosteuropas. Langfristig betrachtet, sichern die mit der Beteiligung verbundenen Synergien das Wachstum der gesamten Unternehmensgruppe der Mobilkom Austria [MoAu05b].

Auch für Mobilkom Austria steht die Unternehmensphilosophie des *Customer Relationship Managements* im Vordergrund. Mobilkom Austria realisierte eines der modernsten *Customer Service Centers* und betreibt unter der Marke *client24* innovative Services für die Bereiche Kundenbetreuung, Kundenbindung und Kundengewinnung [MoAu05b].

4.2.2 Optimierung der Web-Dienste

Die Umsetzung der Einflussfaktoren durch den Einsatz der vorgestellten informationstechnologischen Instrumente ermöglicht die Optimierung der Web-Dienste. Einerseits erleichtert die Anwendung des *Semantic Web Mining* die relevante Informationsgewinnung und andererseits stiftet das *Recommender System* einen Mehrwert indem es auf Kundenprobleme aufmerksam macht und Entscheidungsfindung seitens des Kunden erleichtert. Das semantische Web stellt die besten Voraussetzungen für den Einsatz dieser Instrumente und für die Optimierung weiterer Web-Dienste dar. Nachfolgend werden die einzelnen Einflussfaktoren in Bezug auf die genannten Ziele untersucht und mit jeweils einem spezifischen Beispiel erläutert. Die Optimierung der Web-Dienste wird für den ersten Einflussfaktor anhand des Beispiels über Suchabfragen im semantischen Web dargestellt. Dabei liefert das maschinenfreundliche Web auch auf offene und wagen Suchabfragen über Mobiltelefone besonders gute Antworten. Für den zweiten Einflussfaktor wird die Optimierung der Web-Dienste an den einfachen, benutzerfreundlichen Webseiten

der One GmbH und Mobilkom Austria präsentiert. Der dritte Einflussfaktor erfordert die Optimierung der individuellen Beziehung zum Kunden mithilfe von Personalisierungsmethoden, wodurch ein kundenfreundliches Web realisiert wird.

1. Einflussfaktor: Technologie für das Web

Wie im Kapitel 4.2.1 gezeigt wurde, ist die erfolgreiche Unternehmensphilosophie *Customer Relationship Management* (CRM) bei den vorgestellten Unternehmen im Einsatz. CRM ist in der Praxis nicht nur sehr verbreitet sondern auch besonders erfolgreich. Daher wird in weiterer Folge davon ausgegangen, dass ein in der Praxis erfolgreiches Unternehmen sich stets auf die Grundkonzepte des CRM orientiert. Um nun die Signifikanz des Einflussfaktors *Technologie* für die Praxis darzustellen, wird CRM herangezogen. Dabei ist CRM ein effektives Instrument zur Messung der Signifikanz informationstechnologischer Instrumente für die Praxis. Die Relevanz des Einflussfaktors für CRM wird anschließend dargestellt.

Customer Relationship Management dient der Intensivierung der Kundenbeziehung. Dabei wird die Beziehung zum Kunden gepflegt um insbesondere das Vertrauen und die Zufriedenheit des Kunden zu erhöhen. Um diese Kundenbeziehung auch im Web bestmöglich zu realisieren, ist primär die Verständnisgrundlage zwischen Mensch und Maschine zu optimieren. Nicht nur der Benutzer soll die Information im Web gut verstehen, sondern auch der Computer soll die Daten und Informationen des Benutzers richtig interpretieren können. Dies ist im *World Wide Web* leider nicht gegeben. Aber durch den Einsatz von Ontologiesprachen kann *Semantic Web* aufgebaut werden, wodurch auch der Computer fähig wäre Informationen richtig zu verstehen und den Benutzer zum Beispiel bei seiner Suche im Web bestmöglich zu unterstützen. Das Web der zweiten Generation sorgt somit für besseres Verständnis zwischen Mensch und Maschine. Das Semantic Web unterstützt die Intensivierung der Kundenbeziehung aufgrund der verbesserten Verständnisgrundlage zwischen Benutzer und Computer. So kann die Produktsuche eines potentiellen Kunden optimiert werden, wodurch er nicht nur treffende und zufrieden stellende Antworten auf seine Suchabfrage erhält sondern dadurch auch Zeit sparen kann.

Das CRM im Web lässt sich durch die Technologie des semantischen Webs optimieren, indem das Unternehmen seine Webseite auf die technologischen Anforderungen des semantischen Webs anpasst. Die Wirkung des CRM lässt sich an den Aktivitäten des Kunden auf der Webseite erkennen. Somit stellt die Webseite ein Mittel zum Zweck dar, obwohl sich in diesem Prozess lediglich zwei Akteure befinden, nämlich der Kunde und das Unternehmen (indem der Kunde die Webseite besucht und das Unternehmen die Webseite betreibt). Da sich das CRM im Web jedoch nur mithilfe der Webseite realisieren lässt, wird die Webseite im Bild 15 als dritter Akteur dargestellt.

Technologische Anforderungen sind für jede kommerzielle Webseite gegeben, insbesondere in der Zeit des fortschreitenden Informationszeitalters und der technologischen Entwicklungen und Innovationen. Das Web der 2. Generation ist eine solche technologische Weiterentwicklung. Die Entwicklung kommerzieller Webseiten in stabilen Ontologieumgebungen wird den Anforderungen des Semantic Web gerecht. Wie bereits anhand der Berners-Lee's Architektur im Kapitel 3.1 gezeigt wurde, kann das semantische Web mit dem Einsatz von Ontologiesprachen aufgebaut werden.

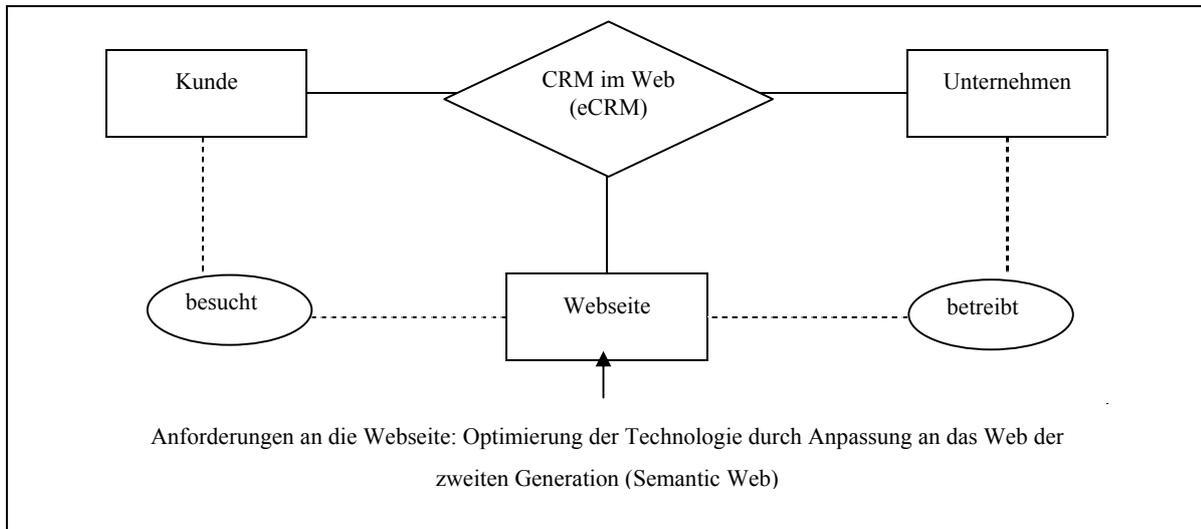


Bild 15 Customer Relationship Management im Web

In der Praxis gibt es somit mehrere Akteure, die im Web zwar zusammenarbeiten aber aus unterschiedlicher Motivation heraus aufeinandertreffen, hauptsächlich zum Zwecke von Absatz oder Unterhaltung. Wird nun ein in der Praxis stattfindender Vorgang im Web verallgemeinert, dann lassen sich insbesondere die Entitäten *Maschine*, *Daten* und *Mensch* abstrahieren und zueinander in Beziehung setzen. Die Maschine bzw. der Computer verarbeitet die Daten (zum Beispiel als eine Reaktion auf eine Benutzer-Abfrage). Diese Verarbeitung ist umso genauer, je besser der Computer die Daten interpretieren kann. Der Mensch speichert eine Vielzahl von Daten und ruft andere Daten wieder ab. Werden nun Ontologiesprachen (zum Beispiel RDF) für die Speicherung von Daten eingesetzt, dann werden dem Computer die besten Voraussetzungen für eine optimale Weiterverarbeitung dieser Daten ermöglicht. Diese Weiterverarbeitung erfolgt in Praxis zum Beispiel im Falle einer Suchabfrage. Bild 16 stellt den beschriebenen Sachverhalt graphisch dar.

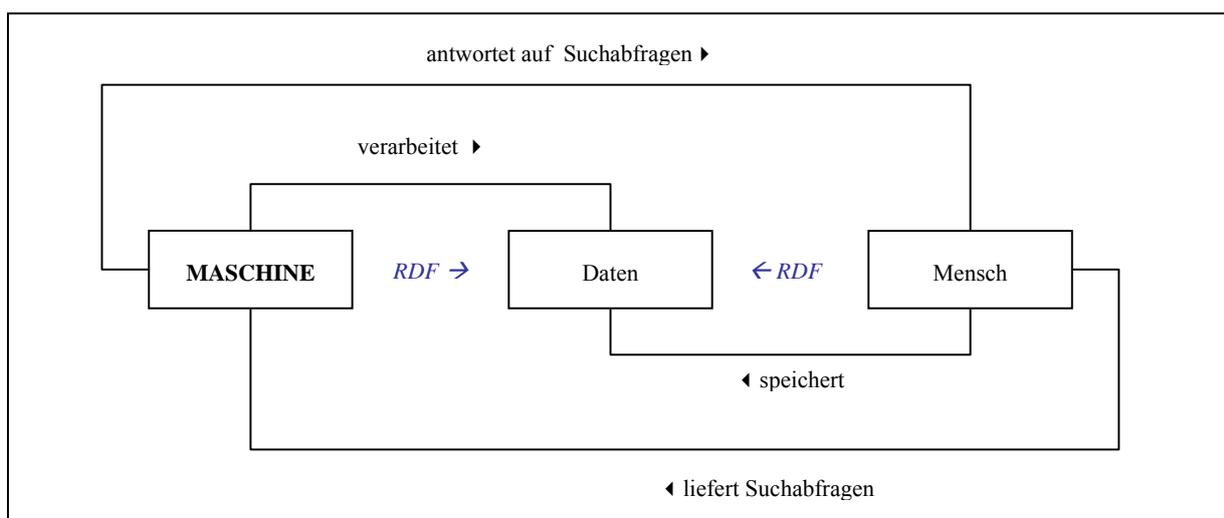


Bild 16 Abstraktion eines Web-Vorgangs (Maschine)

Steht nun die Maschine im Zentrum der Überlegungen, wird der Nutzen für die verschiedenen Akteure Berücksichtigung finden. Der Anwender (zum Beispiel der Kunde) profitiert von der rasch beantworteten Suchabfrage. Anhand des folgenden Beispiels wird die Funktionsweise der Informationsgenerierung im semantischen Web erklärt.

Beispiel:

Zum Zweck der Anschaffung eines Telekommunikationsgerätes sucht der Kunde nach relevanten Informationen im semantischen Web. Seine Abfragen hinsichtlich eines *Smart phones* sind wage, aber seine Vorstellung umso vielfältiger.

Abfrage 1: WAP-fähiges Handy mit Kamera von hoher Bildqualität

Der Kunde weiß nicht ab wie viel Pixel eine bestimmte Bildqualität bei Kameras gewährleistet ist und er weiß auch nicht welche Hersteller solche Handies produzieren. Im semantischen Web würde die Suchmaschine mit „hoher Bildqualität“ den Hersteller „Samsung“ und „1-Megapixel“ assoziieren und auch direkt auf entsprechenden Webseiten nach dem Produkt suchen. Mit derselben Abfrage würde in Anbetracht des technologischen Fortschritts die Suchmaschine in fünf Jahren andere Parameter assoziieren. So wäre es durchaus vorstellbar, dass die Suchmaschine mit hoher Bildqualität in Zukunft „10-Megapixel“ und einen heute noch unbekanntem Hersteller assoziiert. Der Kunde stellt noch eine Abfrage an die Suchmaschine:

Abfrage 2: WAP-fähiges Handy mit MP3-Player und Radioempfang zum Studententarif Mit der Einschränkung: Suche nur Seiten aus Inland

Der Kunde weiß nicht ob ein Netzanbieter im Inland ein solches Handy anbietet. Er hat nur davon gehört und er ist auch unentschlossen hinsichtlich der Wahl des Netzanbieters. Mit Studententarif assoziiert die Suchmaschine einen „Netzbetreiber“ und schließt daher die Webseiten der Hersteller und anderer Anbieter aus. Mit der Einschränkung „nur Seiten aus Inland“ werden lediglich kommerzielle Seiten von Unternehmen aus Österreich berücksichtigt und die aktuellen Angebote selektiert. Mit „Inland“ assoziiert die Suchmaschine jenes Land, in dem der Kunde seinen Internetprovider hat. Der Benutzer kann daher im Semantic Web offenere Abfragen stellen und trotzdem sehr rasch eine passende Antwort erhalten. Das Unternehmen kann sich dadurch sicher sein, dass potentielle Kunden rascher an die gewünschten Informationen oder angebotenen Produkte gelangen.

2. Einflussfaktor: Information aus dem Web

Um den Anforderungen des CRM im Web (eCRM) gerecht zu werden, ist die Information über den Kunden welche in Form von hinterlassenen Spuren auftreten, zu speichern und zu analysieren. Da das Web Mining das Verhalten von Benutzern misst und analysiert, eignet sich das Instrumentarium zur erfolgreichen Verkaufsmustererkennung sowie das *Semantic Web Mining* zur Verkaufsprozessoptimierung. Dadurch lässt sich der Umsatz einer E-Business Anwendung effektiv erhöhen [RoVo02]. Durch die Analyse der Informationen über Kunden lassen sich für das CRM nützliche

Deduktionen ableiten, welche zur Lösung unterschiedlicher Problemstellungen herangezogen werden können. Im Zentrum der Überlegung stehen nun die Daten bzw. die Inhalte einer kommerziellen Webseite. Daher ist die Entität *Daten* im Bild 17 mit Großbuchstaben hervorgehoben.

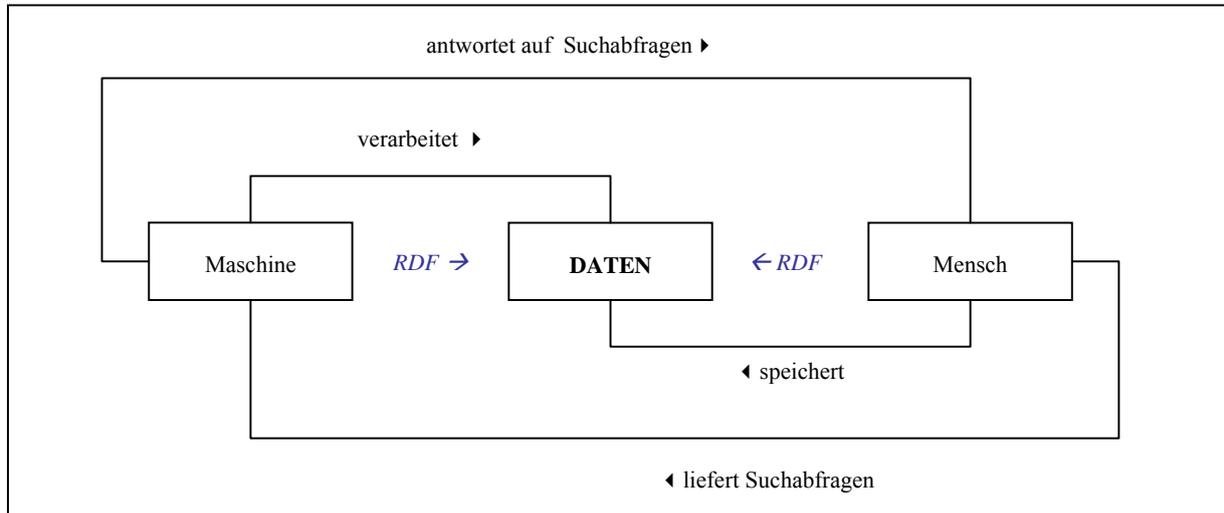


Bild 17 Abstraktion eines Web-Vorgangs (Daten)

Die Sicherung der Profitabilität eines Online-Geschäftsmodells erfordern relevante Informationen über Benutzersegmente sowie die von Benutzern ausgelösten Transaktionen. Die dabei in Mittelpunkt stehenden Fragen, sollen anhand von *Semantic Web Minig* beantwortet werden. In der Literatur [RoVo02] werden folgende Beispiele solcher Fragen präsentiert:

- Wie viele Besucher kommen wann in den Shop?
- Wer sind meine Besucher – wo und wie werden sie zu profitablen Kunden?
- Welche Wege führen am häufigsten zur Bestellung?
- Welche Produkte und Themen sind wann besonders beliebt?
- Wovon hängt der zeitliche Verlauf der Bestellungen ab?
- Wofür wird die Such-Funktion genutzt und wann wird „Hilfe“ benötigt?
- Wann und wo werden häufig Bestellungen abgebrochen?
- Wie optimiere ich das Sortiment?

Die Menge der notwendigen Daten und Informationen steigt mit der Zunahme der Problemkomplexität [Scha03]. Die Anwendung des Web Mining ist als ein Konzept für die Bündelung einer Reihe von einzelnen Web Mining Dienstleistungen zu verstehen [RoVo02].

Laut Roth et al. [RoVo02] bestimmen folgende drei Dimensionen die erfolgreichen Online-Shopping-Anwendungen:

- Technik (Funktionalitäten, Leistung, Stabilität, Sicherheit etc.)
- Angebot (Inhalt, Sortiment, Navigation, Design, Benutzerfreundlichkeit etc.)
- Nachfrage (Erfassung, Analyse, Bewertung, etc.)

Die besten Datenanalysen bleiben unwirksam, wenn keine Umsetzung des Erkenntnisgewinns der Analysen im Geschäftsprozess und im Shopsystem erfolgt. Das Potential des Web Mining lässt sich durch die *aufeinander abgestimmten Serviceleistungen* ausschöpfen. Ökonomische Gründe sprechen für die Bereitstellung von Anwendungsdiensten, sog. *Application Service Providing* (ASP) [RoVo02].

Das vorgestellte Konzept des *Semantic Web Mining* stellt wesentliche Innovation dar, durch welche Informationen schneller und effektiver auffindbar werden. Das *Semantic Web Mining* setzt aber semantisches Web voraus, das heißt die Verwendung semantischer Strukturen und Ontologiesprachen. In Verbindung mit der digitalen Wertschöpfung kommt der Webtechnologie besondere wirtschaftliche Bedeutung zu – nicht nur für einzelne online-operierende Unternehmen, sondern für das gesamte *Electronic Business* (Definition des Begriffs im Kapitel 3.1).

Die Voraussetzung für erfolgreiche Anwendungen des Webs stellt eine performante technische Systemlösung mit benutzerfreundlicher Bedienbarkeit, übersichtlichen Design und kurzen Wegen. Dazu gehören auch minimale Ladezeiten der Seiten, schnelle Antwortzeiten der Server sowie Stabilität und Sicherheit, die durch erstklassige technische Lösungen gewährleistet werden sollen [RoVo02]. In der Praxis können die kommerziellen Webseiten in vielfacher Weise optimiert werden.

Die Art wie sich das Unternehmen präsentiert und welche Webleistungen es anbietet gehören zu den Kriterien der erfolgreichen Webpräsentation eines Unternehmens und werden daher im folgenden Beispiel zum Gegenstand. Dabei werden die Webseiten von One und Mobilkom Austria hinsichtlich ihres Aufbaus und der Benutzerfreundlichkeit untersucht.

Beispiel:

In Kapitel 3.1 wurden einige Beispiele für Dimensionen einer internationalen Webseite präsentiert. An dieser Stelle wird vom Grundgedanken her äquivalent vorgegangen. Die Mobilkom Austria Webseite (<http://www.a1.net>) lässt sich anhand folgender Dimensionen darstellen: *Privat*, *Business* und *Online Shop*. Jede dieser Dimensionen weist bereichsspezifische Untermenüs auf. Dieser Aufbau der Webseite bietet insbesondere gute Übersichtlichkeit und einfache Navigation für den Benutzer. Zudem sind die Hauptnavigationsleisten durch Nebennavigationsleisten ergänzt, was zur Optimierung der Übersichtlichkeit führt [MoAu05a]. Auf der Webseite des Netzproviders One lassen sich folgende Dimensionen zum 2006-04-18 differenzieren: *Privatkunden*, *Wertkartenkunden*, *Businesskunden* und *Kundenservice*. Die einfache Strukturierung und klarer Überblick der Inhalte und Informationen lassen die Webseiten insbesondere benutzerfreundlich wirken, was letztendlich das Ziel des Einflussfaktors *Information* darstellt.

Da es für die Optimierung von Webseiten noch weitreichende Methoden in der Literatur gibt, werden diese im eigenen Kapitel 4.2.3 behandelt. Dabei werden drei Verfahren von Bensberg zur Optimierung von kommerziellen Webseiten dargestellt.

3. Einflussfaktor: Personalisierung im Web

“To give customers exactly what they want, you first have to learn what that is. It sounds simple, but it’s not.” [Zipk01]

Im Zusammenhang mit der *Personalisierung im Web* und mit der Unternehmensphilosophie des *Customer Relationship Managements* hat sich der Begriff Kundeninteraktion etabliert. Die Kundeninteraktion dient der Verstärkung bzw. Vertiefung der Kundenbeziehung und sorgt für erfolgreiche Kundengewinnung. Das Ziel der Kundeninteraktion besteht in der Sammlung von geeigneten Daten, Informationen und Wissen über den Kunden. Auf Basis relevanter Informationen können Kundenbedürfnisse abgeleitet werden und nachfragende, kundenspezifische Güter zur Befriedigung dieser Bedürfnisse identifiziert werden [Scha03]. Die Datenaufbereitung und sinnvolle Analysen sind für die *Personalisierung im Web* unentbehrlich.

Im Zentrum der Überlegungen steht nun der Kunde bzw. ein Anwender des Webs. Der Mensch speichert nicht nur die Daten sondern ruft andere Daten aus beruflichen oder privaten Zwecken auf und verlässt sich auf die zum Einsatz kommende Technologie. Im Bild 18 ist daher die Entität *Mensch* mit Großbuchstaben hervorgehoben.

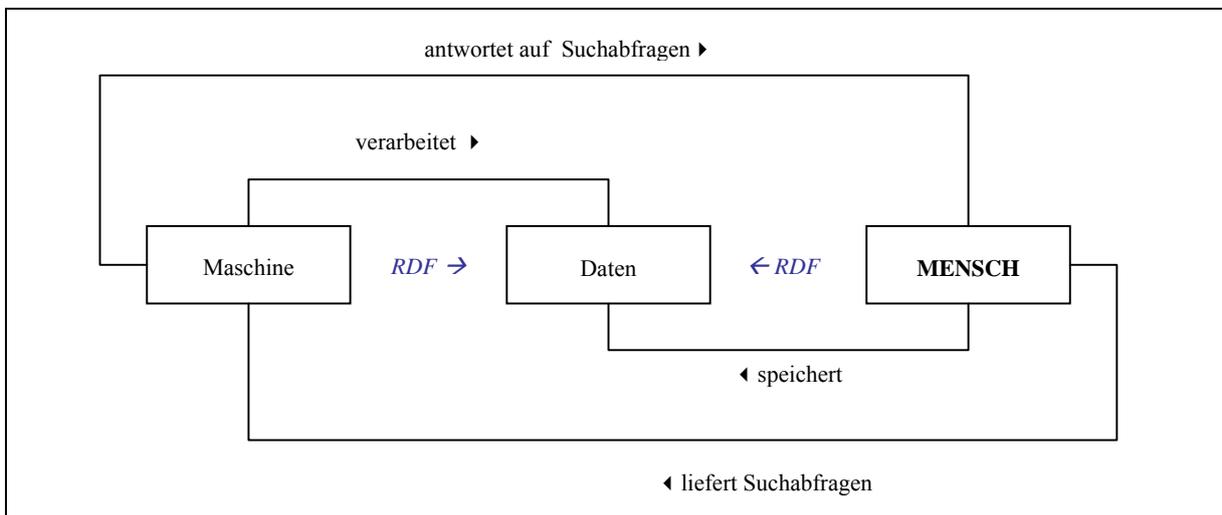


Bild 18 Abstraktion eines Web-Vorgangs (Mensch)

Einige Begriffe wie Kundenproblem, Problemlösungen, Problemlösungsprozess sowie die Zusammenhänge zwischen den Begriffen sollen in Folge erläutert werden. Das Kundenproblem stellt den *Auslöser* für die Erstellung eines individuellen Guts dar. Der Kunde kann sich selbst seines Problems

bewusst werden und darauf reagieren oder er kann auf das Problem aufmerksam gemacht werden [Scha03].

Marktfähige Produkte können aus der Sicht des Kunden als eine Art von *Problemlösungen* interpretiert werden [ReMö97]. Die Bedürfnisse der Kunden stellen die Probleme dar, die das Unternehmen durch geeignetes Angebot zu lösen versucht. Demnach sind Güter die *Problemlösungen* aus der Sicht des Kunden [ReMö97]. Das Angebotssortiment ist den Problemlösungen gleichzustellen.

Das Kundenproblem initialisiert einen *Problemlösungsprozess*, im Rahmen dessen die Leistungsspezifikation erfolgt. Die Leistungsspezifikation beschreibt die Lösung des Kundenproblems anhand von Leistungsmerkmalen [Scha03]. Erst die Leistungserbringung stellt die Lösung des Problems dar. In dem die Leistung letztendlich in den Verfügungsbereich des Kunden gelangt, wird es dem Kunden erst dann möglich, diese Leistung zur Problemlösung einzusetzen, das heißt zu konsumieren [Scha03].

Im Zusammenhang mit dem Problemlösungsprozess ist auch das *Problemlösungswissen* notwendig, durch dieses erst das Kundenproblem richtig erkannt und beschrieben werden kann. Das heißt das Wissen darüber, dass ein Problem vorhanden ist und dass Problemlösungen existieren und von bestimmten Anbietern zur Verfügung gestellt werden können. In weiterer Folge lassen sich durch die notwendigen Informationen über das Problem entsprechende Leistungsspezifikationen ableiten [Scha03].

Zudem ist auch das Wissen über die relevanten Instrumente der Informationstechnologie wie zum Beispiel *Semantic Web Mining* ausschlaggebend, wodurch sich das Kundenproblem identifizieren lässt. Das *Recommender System* kann zudem herangezogen werden um den Kunden an bestimmte Problemlösungen aufmerksam zu machen.

Das Bedürfnis des einzelnen Kunden ist der Auslöser für das personalisierte Angebot des Unternehmens. Die individualisierte Nachfrage stellt das Kundenproblem dar, welches durch entsprechende Problemlösung vom Unternehmen gelöst werden kann. Auf Basis des Kundenproblems wird die Problemlösung erarbeitet und das individuelle Gut angeboten.

Im Zusammenhang mit der Personalisierung im Web tragen insbesondere kundenspezifische Web-Leistungen zur Optimierung von Web-Diensten bei. Der Kunde reagiert auf persönlich an ihn gerichtete Informationen und Empfehlungen in Abhängigkeit von seinem Interessensgebiet. Empfehlungen über Angebote, die ihn nicht interessieren, sind auf alle Fälle nutzlos für ihn. Andere Empfehlungen, die sein Interesse sogar wecken, generieren ihm einen Mehrwert. Die Personalisierung erfordert daher die Anpassung der Inhalte an die Interessen des einzelnen Kunden.

Beispiel:

Im Speziellen bei Vertrieb von Mobiltelefonen und damit in Verbindung stehenden Funktionalitäten (GPRS, UMTS, EDGE, WAP etc.) gibt es verschiedene Möglichkeiten zur Personalisierung. Beispiele personalisierter Inhalte finden sich in Form von explizit vom Kunden erwünschten Mitteilungen aufs Handy, individuelle bzw. kundenspezifische Anwendungen auf der Webseite des Netzbetreibers sowie zusätzliche und vielfältige Dienstleistungen rund ums Handy.

Die tatsächliche Durchführung einer Bestellung oder beendete Auftragsabwicklung ist insbesondere von der Webseite bzw. von der Qualität der Webleistungen abhängig. Die

Mehrwertangebote auf Webseiten ermöglichen personalisierte, auf Kunden zugeschnittene Seiten, welche sich insbesondere durch personalisierte Empfehlungssysteme optimieren lassen. Mit Zunahme der kundenspezifischen Mehrwertangebote steigt aus der Sicht des Käufers auch die Qualität der Webseite. Die Qualität ist zudem von der Webtechnologie bzw. von der Forschung und Entwicklung in diesem Bereich geprägt.

4.2.3 Optimierung der kommerziellen Webseiten

Jedes Unternehmen, das online operiert, präsentiert sich durch seine Webseite. Aufgrund dieser Präsentation kommt dem Aufbau der Webseite wichtige Bedeutung zu. Das Ziel liegt insbesondere darin, die Webseite benutzerfreundlich zu gestalten indem die Webseite für den Benutzer möglichst einfach aufgebaut wird. Die Verfahren von Bensberg zeigen wie insbesondere kommerzielle Webseiten hinsichtlich ihrer Benutzerfreundlichkeit optimiert werden können.

Verfahren zur Webseiten-Optimierung nach Bensberg

Das Web Mining bietet ein umfangreiches, methodisches Instrumentarium, das eine Informationsgrundlage zur nachhaltigen Optimierung von Webseiten schaffen kann [Bens02].

Sinkende Kontaktzahlen, kurze Besuchsdauer oder stagnierende Umsätze in Online-Shops machen den Einsatz von Optimierungsmaßnahmen der Webseite erforderlich. Dabei differenziert Bensberg [Bens02] zwischen unterschiedlichen Interaktionsmechanismen:

- Der Online-Besucher navigiert im Rahmen des Browsing im Informationsangebot einer Webseite. Dieses Informationsangebot entspricht elektronischem Produktsortiment, das auf der Webseite hierarchisch dargestellt ist.
- Der Online-Besucher hat die Möglichkeit, mittels Textsuche gezielt nach bestimmten Begriffen zu suchen. Die Suchergebnisse werden dabei über Listen, welche die Verweise auf weiterführende Informationen enthalten, visualisiert.

Des Weiteren unterscheidet Bensberg [Bens02] zwischen internen und externen Merkmalen einer Webseite:

- Interne Merkmale einer Webseite wie zum Beispiel Navigationsmechanismen eines Online-Shops können bei Bedarf Gegenstand des Optimierungsverfahrens sein.
- Externe Merkmale sind zum Beispiel Suchmaschineneinträge einer Webseite, die ebenfalls optimiert werden können.

Durch die externe Optimierung der Webseite kann der Auftritt am elektronischen Markt verbessert bzw. intensiviert werden. Demgegenüber wird mit der internen Optimierung die kundenorientierte Ausgestaltung von Online-Shops verbessert. Dadurch lässt sich die Kaufneigung von Online-Besuchern steigern [Bens02]. Bild 19 zeigt den Anwendungsbereich der internen und externen Mechanismen.

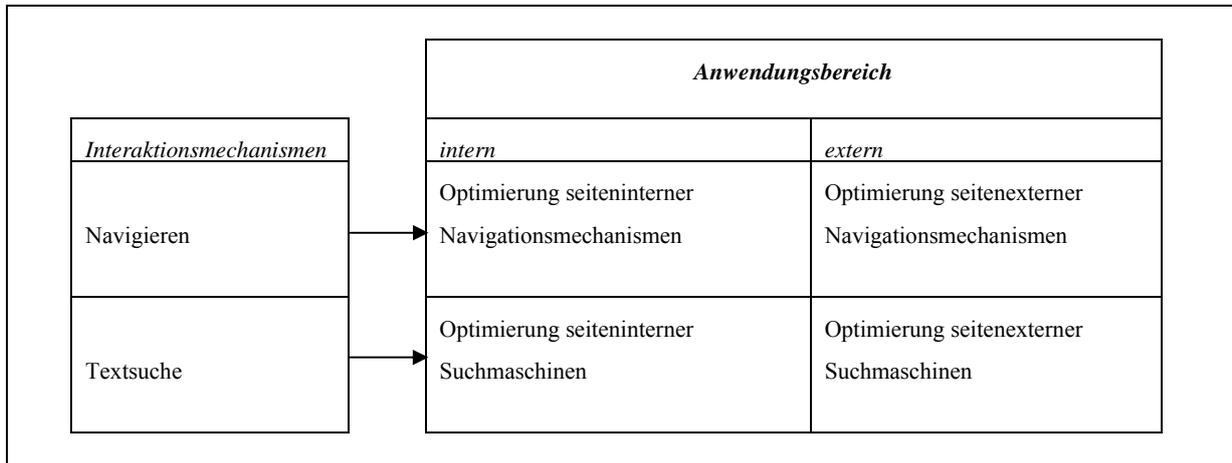


Bild 19 Maßnahmen der Webseiten-Optimierung [Bens02]

Laut Bensberg [Bens02] lassen sich anhand der seiten-interner Interaktionsmechanismen detaillierte Analysen des Interaktionsverhaltens der Online-Kunden erstellen und auf diese Weise auch folgende Fragestellungen beantworten:

- Welche Informationsangebote und Produkte werden häufig gemeinsam gesucht bzw. gekauft?
- Welche Pfade nutzen die Online-Kunden am häufigsten?
- Mit welchem Erfolg werden Online-Geschäftsprozesse durchgeführt?

Bensberg [Bens02] stellt Optimierungsverfahren vor, die der Beantwortung der oben genannten Fragen dienen. Dabei unterscheidet er zwischen *Optimierungsverfahren auf Basis von Verbundbeziehungen*, *auf Basis von Pfadanalysen* und *auf Basis von Geschäftsprozessen*. Im Folgenden werden diese Verfahren kurz vorgestellt.

Optimierung auf Basis von Verbundbeziehungen

Zwischen Informationsangeboten und Produkten lassen sich Verbundbeziehungen identifizieren. Auf Basis dieser Beziehungen zwischen Produkten und Angeboten (Gemeinsamkeiten, Ähnlichkeitsgrad, komplementäre Güter etc.) lassen sich Produkte in Online-Shops gemeinsam präsentieren [Bens02].

Optimierung auf Basis von Pfadanalysen

Die zeitlich-sequentielle Analyse liefert Aussagen über die Nutzung bestimmter Pfade. Mit den Sequenz- bzw. Pfadanalysen lässt sich beispielsweise die Fragestellung beantworten, welche Informationsangebote die Kunden nach Abruf der Standardseite wählen. Dadurch lässt sich eruieren, welche Informations- und Produktangebote für die Online-Besucher nach Betreten der Webseite einen hohen Interessantheitsgrad aufweisen. Anhand der Pfadanalyse können insbesondere jene Pfade identifiziert werden, die signifikante Navigationszusammenhänge kennzeichnen. Diese Evaluation der Informationsangebote hinsichtlich ihrer Verhaltensrelevanz bildet die Grundlage für die Integration jener Navigationsmechanismen, welche die Präferenzen der Besucher reflektieren [Bens02].

Optimierung auf Basis der Geschäftsprozesse

Zur Optimierung von Webseiten ist eine eingehende Analyse unternehmenskritischer Geschäftsprozesse notwendig, die in der Praxis Aufschluss über die einzelnen Prozessmengen und Prozessdurchlaufzeiten, liefert. Beispiele dafür sind insbesondere Bestell- und Buchungsprozesse, anhand derer sich teilweise der ökonomische Erfolg einer Webseite ableiten lässt [Bens02].

Die Geschäftsprozessanalyse wird in [Bens02] anhand eines Bestellprozesses beschrieben. An dem Bestellprozess nehmen 20.000 potentielle Käufer teil, die im Durchschnitt 15 Minuten für die Produktsuche und Befüllen des Warenkorbs benötigen. Von den 20.000 Besuchern, überlegen sich 12.000 Besucher auf Grund von internen oder externen Faktoren anders und leeren den Warenkorb wieder. Die übrigen 8.000 Besucher lösen den Bestellprozess aus, jedoch erfolgen die Adressdateneingabe und die Wahl der Zahlungsbedingungen nicht immer wie gewünscht (aufgrund von technischen oder launischen Komponenten). So dass schließlich nur 2.304 Besucher die Bestellung tatsächlich abschließen.

Tabelle 3 zu Geschäftsprozessanalyse zeigt abschließend, dass insgesamt nur 11,52% von 20.000 potentiellen Käufern die ausgesuchten Güter, die sie in den Warenkorb gelegt haben auch tatsächlich bestellt haben. Kritische Prozessphasen werden im Laufe der Geschäftsprozessanalyse durch simultane Analyse der phasenspezifischen Besuchsdauern und der Konversionsraten identifiziert [Bens02].

Prozessschwachstellen sind in der Phase *Zahlungsbedingungen wählen* beispielsweise die relativ lange Besuchsdauer und gleichzeitig niedrige Konversionsrate. Die Anpassung der formalen oder inhaltlichen Ausgestaltung kann die Anzahl der abgeschlossenen Bestellungen erhöhen [Bens02].

Die Konversionsrate gibt im Zusammenhang mit Geschäftsprozessen das Verhältnis zwischen der Anzahl an Webseiten-Besuchern und den Käufern an. Vielfach wird aber die Gesamtanzahl der Besucher zu optimieren versucht um sich dadurch zumindest eine bestimmte Anzahl der Kunden zu sichern. Denn man kann davon ausgehen, dass durch den Anstieg an Besuchern auf der Webseite neue Kunden dazukommen, während gleichzeitig bestehende Kunden eventuell zu Konkurrenten wechseln.

<i>Prozesse:</i>	Produktsuche, Warenkorb befüllen	Bestellprozess auslösen	Adressdaten eingeben	Zahlungs- bedingungen wählen	Bestellung abschließen
<i>Prozessmenge (Besucher):</i>	20.000	8.000	4.800	3.840	2.304
<i>Durchschnittliche Prozessdauer:</i>	15 Min	1 Min	2 Min	3 Min	1 Min
	40%				
	60%				
	80%				
				60%	
	11,52%				

Tabelle 3 Geschäftsprozessanalyse am Beispiel des Bestellprozesses [Bens02]

Bisherige innovative Visionen im Bereich der Informationswirtschaft und der Informationstechnologie haben sich aufgrund der Beliebtheit des World Wide Web erfolgreich durchgesetzt, sowohl im Bereich der Software- als auch Hardwarekomponenten. Diese Beliebtheit sorgt auch für die Akzeptanz neuer informationstechnologischer Instrumente, wenn diese die Web-Applikationen für den Anwender vereinfachen und für den Anbieter hinsichtlich ihrer Funktionalitäten optimieren. Die Berücksichtigung der genannten Einflussfaktoren setzt den Einsatz von empfohlenen informationstechnologischen Instrumenten voraus, welche wiederum positiven Einfluss auf wirtschaftliche Aktivitäten im Web ausüben. Die wirtschaftliche Bedeutung der präsentierten Einflussfaktoren liegt daher in deren ökonomischen Anwendung, die sowohl dem Anbieter als auch dem Anwender entsprechenden Nutzen stiftet. Der Anbieter von Web-Diensten ist stets ein online-tätiges Unternehmen. Der Anwender dagegen kann ein Web-Besucher, ein System-Benutzer oder ein Kunde sein. Durch die Anwendung der informationstechnologischen Instrumente lassen sich die bereits genannten Ziele erreichen, wodurch das Web maschinen-, benutzer- und kundenfreundlicher wird. In diesen Zielen steckt ökonomisch gesehen hohes wirtschaftliches Potential sowohl aus der Sicht des Anwenders als auch aus der Sicht des Anbieters.

Die vorgestellten Einflussfaktoren stiften für verschiedene Akteure entsprechenden Nutzen. Daraus lassen sich diverse Perspektiven der Betrachtung erkennen. Wie bereits in der Einführung (Kapitel 1) angesprochen, stehen den Anbietern der Informationstechnologie die Anwender gegenüber. Die Anbieter sind daran interessiert die wirtschaftlichen Aktivitäten im Web zu optimieren, indem sie zur Verbesserung der Technologie beitragen, relevante Information filtern und die Personalisierung im Web realisieren. Das primäre Ziel liegt in der Optimierung der eigenen Leistung. Die Anwender der angebotenen Informationstechnologie erkennen diese Verbesserung insbesondere in der benutzerfreundlicheren Handhabung, wodurch sich ihre Produkt-Wünsche einfacher umsetzen lassen.

4.3 Signifikanz der Einflussfaktoren für Anwender

Die vorgestellten Instrumente des *Customer Relationship Managements* erleichtern dem Anwender die Suche nach Produkten und helfen bei der Entscheidungsfindung. In jedem Bereich des Lebens, ob im Beruf, im Bildungswesen oder im privaten Bereich, kann der Benutzer im Web unterwegs sein und von den Vorteilen der eingesetzten Instrumente profitieren. Die Instrumente lassen sich in jedem Bereich effizient einsetzen um dem Benutzer im Web entgegenzukommen, die entsprechenden Anwendungen zu vereinfachen und ihn zu motivieren die Webseite erneut aufzusuchen.

Die Inhalte und Produkte einer Webseite wecken das Interesse des aktiven Benutzers, die Motivation und der entsprechende Mehrwert bilden jedoch den Grundstein für andauernde kommerzielle Beziehung. Der Anwender kann allenfalls von der Umsetzung der informationstechnologischen Instrumente profitieren, was sich insbesondere darin bemerkbar macht, dass zufriedene Webseitenbesucher früher oder später zu Kunden werden, auch wenn sie nur gelegentlich das Bestellservice über Web in Anspruch nehmen. Bereits bestehende Kunden welche mit dem Web-Angebot überaus zufrieden sind, können durch häufigere Bestellungen über Web früher oder später zu treuen Kunden werden.

Anhand der Ursache-Wirkung Beziehung lässt sich die Bedeutung für den Anwender der eingesetzten CRM-Instrumente veranschaulichen. Das Interesse des aktiven Benutzers wird generell durch die Inhalte und angebotenen Produkte der Webseite verursacht, wodurch die Wirkung dessen das notwendige Interesse erzeugt, die entsprechende Webseite genauer unter die Lupe zu nehmen. Zusätzliche Dienste wie der Einsatz von Recommender Systemen (= CRM-Instrument) bilden für den Benutzer einen Mehrwert. Der dadurch verursachte Zufriedenheitsgrad wirkt auf die Motivation des Besuchers, die Webseite bei Gelegenheit erneut zu besuchen. So können gelegentliche Besucher zu häufigeren Besuchern werden oder Kunden zu treuen Kunden gemacht werden. Bild 20 stellt die Ursache-Wirkung Beziehung graphisch dar, die anhand des nachfolgenden Beispiels sinngemäß interpretiert wird.

Beispiel:

Der aktive Benutzer (X) einer kommerziellen Webseite stößt auf interessante Inhalte einer Webseite (A) über neueste Modelle von mobilen Endbenutzergeräten. Dieser Inhalt bzw. die Produkte dieser Webseite steigern sein Interesse (B), was ihn dazubewegt die Inhalte der Webseite genauer zu studieren. Aufgrund seiner Aktivität auf der Webseite werden Daten in Form von Bewegungspfaden sowie Suchbegriffe gespeichert und bei erfolgreicher Registrierung auch ein Profil erstellt. Diese gesammelten Informationen über den Benutzer werden von CRM-Instrumenten (C) ausgewertet indem sie entweder zum Zwecke von Trendanalysen eingesetzt oder zur Generierung von Empfehlungen analysiert werden. Die dabei angebotenen Dienste wie zum Beispiel Empfehlungen hinsichtlich interessanter Produkte stiften dem Benutzer einen Mehrwert (D). Durch diese Mehrwertdienste kann der Zufriedenheitsgrad (E) gesteigert werden, was sich wiederum auf die Motivation des Benutzers (F), die Webseite erneut zu besuchen, positiv auswirkt.

Um das Beispiel richtig nachvollziehen zu können, sind im Bild 20 die einzelnen Bestandteile der Ursache-Wirkung Beziehung mit Großbuchstaben versehen.

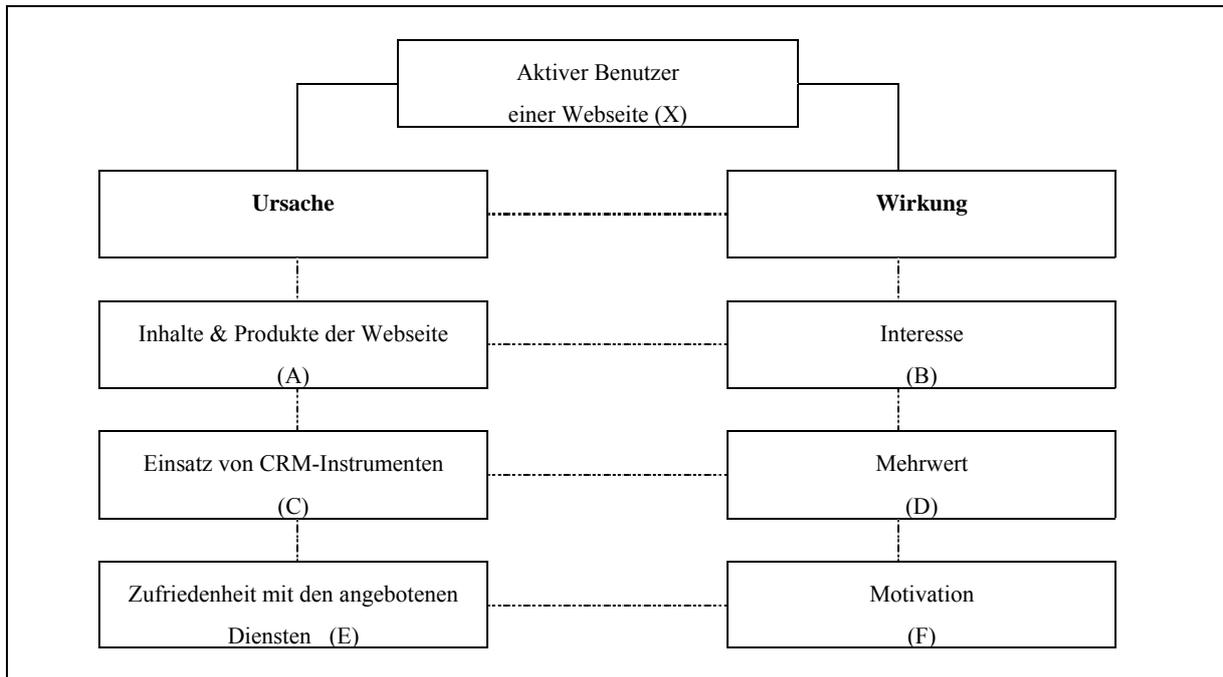


Bild 20 Ursache-Wirkung Beziehung

In Ergänzung dazu wird nun ein kurzes Beispiel gebracht, das im Speziellen auf die Vorzüge des Semantic Web eingeht.

4.3.1 Semantic Web

Im Informationszeitalter von Heute nimmt insbesondere die Web-Technologie die bedeutende Rolle in der technologischen Forschung und Entwicklung (F&E) ein. Der weitere Fortschritt des Informationszeitalters hängt von den Visionen und Ideen sowie von ihrer Realisierung ab. Die Vision von Semantic Web ermöglicht ein verbessertes Web, ein Web der nächsten Generation.

Aus der Sicht des Anwenders ergeben sich zahlreiche Anwendungsmöglichkeiten, welche insbesondere im semantischen Web optimiert werden können. Die Anwendungen erstrecken sich von der Suche nach bestimmten Informationen (zwecks Bildung oder Produktanschaffung), über Downloads aktueller (Open-) Software-Versionen bis zu Abfragen in digitalen Bibliotheken oder Online-Wörterbüchern. Im Zusammenhang mit mobilen Endbenutzergeräten können zusätzliche Web-Dienste sowohl zeitlich als auch räumlich flexibler in Anspruch genommen werden. Das semantische Web würde den Flexibilitätsgrad der mobilen Endbenutzergeräte erhöhen. Insbesondere die zeitliche Komponente würde bei jeder Art von Suchabfragen im semantischen Web optimiert werden.

Beispiel:

Nach einer Suchabfrage im Web bekommt der Anwender eine Antwort, die ihn direkt zur Webseite von Mobilkom Austria führt. Der Besucher erkundigt sich nun nach aktuellen Angeboten, indem er entweder in das vorgesehene Suchfeld den Hersteller des Handys eingibt oder in der Navigationsleiste auf Produkte klickt. Letzteres ermöglicht ihm eine manuelle Suche nach Produktherstellern. Zusätzlich zu der einfachen Suche (mit einem Suchfeld) kann der Besucher mit der erweiterten Suche über verschiedene Parameter (MMS, WAP, GPRS, UMTS etc.) suchen. Im semantischen Web wäre diese Suche nicht notwendig, wenn die semantische Suchmaschine das ideale Angebot sofort anhand der eingegebenen Suchparameter findet. Die Wahrscheinlichkeit für die exakte und zutreffende Antwort auf eine Suchabfrage steigt mit der zunehmenden Semantisierung des Webs.

4.4 Signifikanz der Einflussfaktoren für Unternehmen

„Viele Unternehmen, selbst große Konzerne und etablierte Markenartikler, schöpfen das Potential ihrer Internetanwendungen nicht vollständig aus. Sie präsentieren zwar ihre Produkte ansprechend, ermöglichen eine intuitive Bestellung und sorgen für eine reibungslose Auftragsabwicklung über ihr Warenwirtschaftssystem – der Besucher und sein Verhalten in Online-Shop sind jedoch für viele Unternehmen noch immer eine unbekannte Größe.“ [RoVo02]

Die wirtschaftliche Betrachtung einer Tätigkeit – sei es im Web oder im Shop – erfordert Unternehmen welche zum Zweck eines Absatzes verschiedene Produkte und Dienstleistungen anbieten. Daher ist die Betrachtung informationstechnologischer Instrumente *aus der Sicht des Unternehmens* die naheliegendste Möglichkeit zur Darstellung dessen wirtschaftlicher Bedeutung. Im Zusammenhang mit *Semantic Web Mining* und *Recommender Systemen* sind nur Tätigkeiten im Web von Relevanz. Das Kapitel 4.4.1 stellt anhand einiger Beispiele die wirtschaftliche Bedeutung von Semantic Web Mining für insbesondere online-tätige Unternehmen aus dem Mobilfunknetz dar. Die Beispiele können auch auf andere kommerzielle Webseiten angewendet werden. Da für diese Arbeit die Webseiten von Mobilkom Austria und One ausgewählt wurden, sind die Beispiele primär an den Webseiten des Mobilfunknetzes gerichtet. Ebenfalls aus der Sicht des Unternehmens wird in Kapitel 4.4.2 die ökonomische Signifikanz von Recommender Systemen untersucht.

4.4.1 Semantic Web Mining

Aus der wirtschaftlichen Perspektive betrachtet, gehört *Semantic Web Mining* zu den Instrumenten der Informationstechnologie, die durch ihren Einsatz ökonomische Vorteile für den Einzelnen, für das Unternehmen und für die Gesellschaft erzielen. Vorteile für das Unternehmen ergeben sich durch die Information über den Kunden, die einen gewissen Maß an Personalisierung erlauben und durch die Erstellung von personalisierten Profilen welche wiederum dem Kunden einen Mehrwert bieten. Zudem lässt sich die Akzeptanz des semantischen Webs in der Gesellschaft durch die dabei erzielte Benutzerfreundlichkeit gewährleisten.

Durch den Einsatz des Instrumentes, *Semantic Web Mining*, auf einer kommerziellen Webseite lässt sich das Such- und Einkaufsverhalten eines Benutzers bestmöglich auswerten. Anhand von semantischen Strukturen lassen sich eindeutiger Informationen ableiten. Das dabei gewonnene Wissen über die Bedürfnisse der Webkunden bietet die beste Grundlage für den wirtschaftlich ökonomischen Einsatz von Personalisierungsstrategien. Eine der erfolgreichsten Personalisierungsmethoden ist die der Personalisierungssoftware, welche in dieser Arbeit vorgestellt wurde, das Recommender System.

Der Benutzer initialisiert den gesamten Vorgang des *Semantic Web Mining*, indem er auf der entsprechenden Webseite aktiv wird. Dabei werden sein Such- und Einkaufsverhalten analysiert und die Informationen an Personalisierungssoftware weitergeleitet. Der beschriebene Pfad ist der indirekte Weg der Informationsgewinnung. Der Benutzer hat gelegentlich auch die Möglichkeit seine Präferenzen über Produkte bzw. seinen Stil oder Geschmack auch direkt an die Personalisierungssoftware zu übermitteln. Im Bild 21 wird der direkte Pfad mit der strichlierten Linie angezeigt.

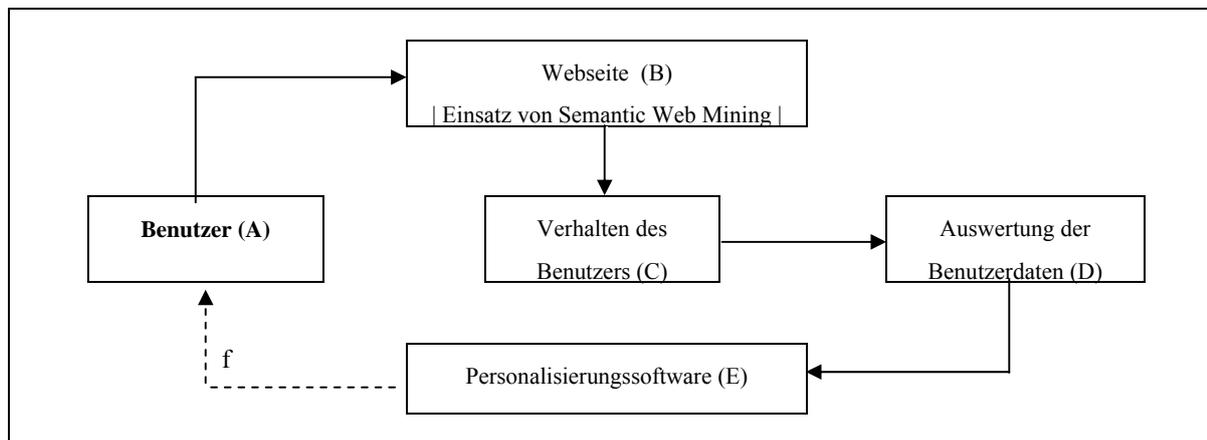


Bild 21 Direkter und indirekter Informationspfad

Beispiel:

Im Bild 21 betritt der Benutzer (A) eine Webseite eines Unternehmens (B) auf der Semantic Web Mining zum Einsatz kommt. Das spezifische Verhalten des Benutzers (C) hinterlässt Spuren bzw. Daten über den Benutzer. Diese Daten lassen sich im nächsten Schritt beispielsweise in Form von Kundenprofilen auswerten (D) und analysieren. Die Personalisierungssoftware (E) kann nun die Profile zum Beispiel auf Ähnlichkeiten zu anderen Benutzern analysieren und Empfehlungen generieren. Der direkte Pfad der Informationsgewinnung (f) setzt Benutzerdaten voraus, welche entweder schon vorhanden sind oder von Benutzer selbst angegeben werden.

Aus dem oben beschriebenen Informationspfad lässt sich die wirtschaftliche Bedeutung des *Semantic Web Mining* für die Beziehung zwischen dem Kunden und Unternehmen ableiten. Das Unternehmen gewinnt durch den Einsatz des Semantic Web Mining relevante Informationen über den Kunden und kann diese zur Pflege der Kundenbeziehung einsetzen. Durch den Einsatz von *Semantic Web Mining* lassen sich daher die Ziele des *Customer Relationship Managements* effizient realisieren. Somit stellt das *Semantic Web Mining* ein hilfreiches Instrument für *Customer Relationship Management* dar. Durch den Einsatz des informationstechnologischen Instrumentes lässt sich daher die Beziehung zum Kunden optimieren.

Beispiel: Mobilfunknetz

Auf der Webseite des *Mobilfunkanbieters One* hinterlässt der Kunde bei jedem Besuch Spuren in Form von aufschlussreichen Daten über die besuchten Seiten, gesuchten Produkte oder andere Inhalte der Webseite, für die der Kunde spezielles Interesse zeigt. Diese gespeicherten Daten können bei wiederholtem Besuch des Kunden mithilfe von Web Mining Methoden analysiert werden. Im semantischen Netz könnte zudem das Semantic Web Mining zum Einsatz kommen, wodurch die Analyse optimiert werden könnte. Insbesondere lassen sich im Falle von ähnlichen Web-Besuchen eines Kunden oder mehrerer Kunden Trendanalysen erstellen mithilfe dessen, das Angebot von *One* hinsichtlich des Trends umgestaltet bzw. angepasst werden kann. Dies gilt für jede Art von Inhalten einer Webseite. Damit das Semantic Web Mining zum Einsatz kommt reicht, wenn jene Webseite auf der es zum Einsatz kommen soll, eine stabile Ontologiemgebung aufweist. Das Unternehmen ist daher nicht auf andere Webseiten aber auch auf keine anderen unternehmensexternen Komponenten angewiesen.

4.4.2 Recommender Systeme

Der Einsatz von Empfehlungssystemen intensiviert die Beziehung zum Kunden, indem Informationen über den Benutzer gesammelt, ausgewertet und personenspezifische Empfehlungen generiert werden. Die Unternehmensphilosophie des *Customer Relationship Managements* (CRM) lässt sich im Web erst durch den Einsatz der genannten Instrumente der Informationstechnologie realisieren. Aus der Perspektive der erfolgreichen Unternehmensphilosophie des CRM sind Instrumente wie *Semantic Web Mining* und

Recommender Systeme die bedeutendsten Werkzeuge im Zeitalter der Informationstechnologie. Aufgrund ihrer ökonomischen Funktion werden diese Instrumente in dieser Arbeit auch als *Instrumente des Customer Relationship Managements* definiert.

Die Recommender Systeme generieren personalisierte Empfehlungen für den Kunden. Somit werden bestimmte Inhalte der Webseite personalisiert, wodurch sich die individuelle Beziehung zum Kunden erkennen lässt. Dabei werden die in Kapitel 2 genannten *Lösungsansätze für Kunden*, die Individualisierung und Personalisierung im Web, verwirklicht.

Durch den Einsatz der informationstechnologischen Instrumente lassen sich auch spezifische Ziele realisieren. Der praxisorientierte Zusammenhang zwischen Instrumenten und ihren Zielen wird in nachfolgendem Beispiel erläutert.

Beispiel:

Die Pflege der Kundenbeziehung (1) ist im Bild 22 als das erste Ziel des CRM dargestellt, welches insbesondere Informationen über den Kunden oder Webseitenbesucher voraussetzt. Die notwendigen Informationen lassen sich durch den Einsatz von Semantic Web Mining (a) sammeln und speichern. Wird zudem auch das Recommender System (b) eingesetzt um die Informationen zu verarbeiten, lässt sich für den Kunden ein sichtbarer Mehrwert erzielen, welches sich auf die Zufriedenheit des Kunden (2) auswirken kann. Somit kann auch das zweite Ziel des CRM realisiert werden.

Bild 22 zeigt definitionsgemäß die Instrumente des CRM und die dazugehörigen Ziele.

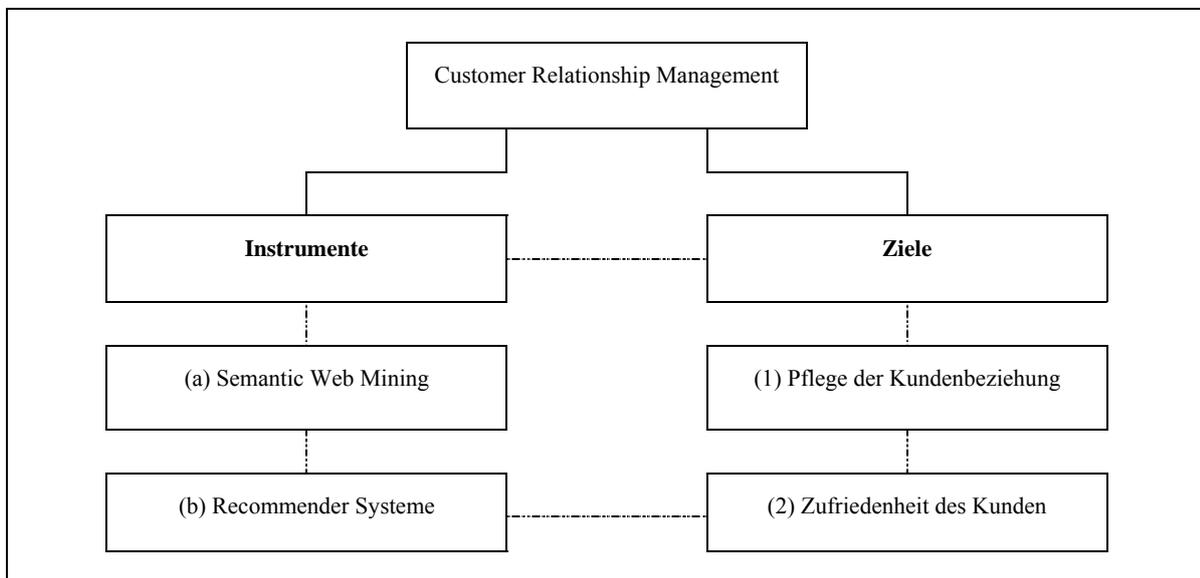


Bild 22 Ziele und Instrumente des CRM

Die wirtschaftliche Bedeutung der beiden Instrumente liegt daher insbesondere in ihrer Funktion als Werkzeuge der erfolgreichen Unternehmensphilosophie jedes gewinnorientierten Unternehmens, das davon Gebrauch macht.

Beispiel: Mobilfunknetz

Der *Mobilfunknetzanbieter Mobilkom Austria* kann die von einzelnen Kunden in Anspruch nehmende Leistungen vergleichen und jene Kunden, die ähnliche Präferenzen (bzw. Geschmack hinsichtlich Produkten und Dienstleistungen) aufweisen in Beziehung setzen um daraus weitreichende Empfehlungen zu generieren. Die angebotenen Empfehlungen stellen Mehrwertangebote dar, die der Kunde unaufgefordert erhält. Die Mehrwertangebote stiften dem Kunden einen Nutzen wenn sie sein Interessensgebiet treffen und ihn zudem auf bestimmte Problemlösungen aufmerksam machen. Je genauer Empfehlungen generiert werden, umso größer ist die Wahrscheinlichkeit den Kunden auf bestimmte Produkte aufmerksam zu machen, welche er in Zukunft auch konsumieren wird.

5 FAZIT

„Forschung und Bildung sind ohne Zweifel das große Kapital post-industrieller Gesellschaften.“ [ReWe06]

Die drei vorgestellten Einflussfaktoren des Webs wurden anhand der Instrumente der Informationstechnologie beschrieben. Der Einsatz dieser Instrumente verbessert die Analyse der vorhandenen Informationen über Kunden (*Semantic Web Mining*) und fördert die personalisierte Beziehung zum Kunden (*Recommender Systeme*). Die in dieser Arbeit vorgestellten Instrumente der Informationstechnologie ergänzen sowohl aus der Anwender- als auch aus der Anbieter-Sicht das Semantic Web.

Der Erfolg eines online-tätigen Unternehmens hängt vom Zufriedenheitsgrad seiner Kunden ab. Daher orientiert sich die wirtschaftliche Bedeutung der informationstechnologischen Instrumente anhand der erfolgreichen Unternehmensphilosophie des *Customer Relationship Managements* (CRM). Jedes erfolgsorientierte Unternehmen fokussiert auf den einzelnen Kunden und nicht auf die Masse. Die Ansätze der Individualisierung und Personalisierung sind daher für CRM wesentliche Komplement-Instrumente. Da die individuelle Beziehung zum Kunden insbesondere im Bereich des Mobilfunknetzes nicht mehr wegzudenken ist, wurden in der praxisorientierten Analyse die Webseiten von Unternehmen der Telekommunikationsbranche (<http://www.a1.net>, <http://www.one.at>) herangezogen.

Das primäre Ziel der Arbeit liegt darin, die wirtschaftliche Bedeutung von informationstechnologischen Instrumenten anhand der identifizierten Einflussfaktoren aufzuzeigen. Die wirtschaftliche Bedeutung der vorgestellten Instrumente für die digitale Wertschöpfung liegt somit in dem Mehrwert, der dem Benutzer durch den Einsatz von Recommender Systemen, geboten wird. Denn jedes online-tätige Unternehmen ist auf den einzelnen Kunden und auf seine Wünsche fokussiert. Die vorgestellten Instrumente, *Semantic Web Mining* und *Recommender Systeme*, dienen insbesondere den Zielen des *Customer Relationship Managements*. Somit lässt sich die wirtschaftliche Bedeutung dieser Instrumente an ihrer nutzenstiftenden Funktion für CRM erklären. Daher werden sie in dieser Arbeit als *Instrumente des CRM* definiert.

Die Intensivierung der Beziehung zum Kunden im semantischen Web erfordert neue Ideen und Visionen, insbesondere im Bereich der Informationstechnologie. Für die neue Generation des Semantic Web heißt es das Web durch semantische Strukturen maschinenfreundlicher zu gestalten, um auf dieser Basis für besseres Verständnis zwischen Mensch und Maschine zu sorgen.

LITERATURVERZEICHNIS

A

[AnHa04]

Antoniou, Grigoris; Harmelen, Frank van: A Semantic Web Primer. Cambridge, Mass 2004.

[ArKo02]

Arndt, Dirf; Koch, Diana: Datenschutz im Web Mining – Rechtliche Aspekte des Umgangs mit Nutzerdaten. In Handbuch Web Mining im Marketing - Konzepte, Systeme, Fallstudien. S.76-103.

B

[BeHL01]

Berners-Lee, Tim; Hendler, James; Lassila, Ora: The Semantic Web. Scientific American, May 2001.

<http://www.scientificamerican.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21&catID=2>, Abruf am 2006-02-04.

[BeHS02]

Berendt, Bettina; Hotho, Andreas; Stumme, Gerd: Towards Semantic Web Mining. Springer, Heidelberg 2002.

<http://www.aifb.uni-karlsruhe.de/WBS/gst/papers/2002/ISWC02.pdf>, 2002, Abruf am 2005-06-04.

[Bens01]

Bensberg, Frank: Web log mining als Instrument der Marketingforschung - ein systemgestaltender Ansatz für internetbasierte Märkte. Dt. Univ.-Verl.; Wiesbaden: Gabler, 2001

[Bens02]

Bensberg, Frank: Website-Optimierung – Aufgabenstellung und Vorgehensweise. In Handbuch Web Mining im Marketing - Konzepte, Systeme, Fallstudien. S.248-265.

[Bern98]

Berners-Lee, Tim: Semantic Web Road Map. <http://www.w3.org/DesignIssues/Semantic.html>. Abruf am 2005-07-25.

[BHHA01]

Bruner, Rick; Harden, Leland; Heyman, Bob; Amato, Mia: NetResult.2 – Best Practices for Web Marketing. New Riders Publishing, 2001.

[Blo01]

Bloehdorn, Stephan: Webmining – Methoden der Ballungsanalyse zur Verdichtung großer Datenmengen, Informationsgewinnung aus dem Internet.

http://www.ira.uka.de/~recht/deu/iir/archiv/2001ss/begleit/inforecht/pass/bloehdorn/bloehdorn_ausarbeitung.pdf, 2001, Abruf am 2005-06-03.

[Bohn03]

Bohnert, Fabian: Einsatz von Collaborative Filtering zur Datenprognose.

<http://www.mathematik.uni-ulm.de/sai/ws03/dm/arbeit/bohnert.pdf>, 2003, Abruf am 2006-04-11.

- [BoWi77]
Bobrow D.G.; Winograd T.: An overview of KRL - a knowledge representation language, Technical Report R76/581 Stanford University, 1977.
- [Böhl92]
Böhler, H.: Marktforschung. 2. Auflage, Stuttgart 1992.
- [Brau05]
Braunschweig, Universitätsbibliothek: Allegro-C – Software für Bibliotheken. <http://www.allegro-c.de/>, 2005-11-17. Abruf am 2005-11-29.
- [Brin03]
Brin, Sergey: Extracting Patterns and Relations from the World Wide Web. <http://www.csee.umbc.edu/%7Ekolari1/Mining/papers/brin.pdf>, 2003-10-24, Abruf am 2005-06-03.
- [BrKH03]
Broekstra, Jeen; Kampman, Arjohn; Harmelen, Frank van: Sesame – An Architecture for Storing and Querying RDF Data and Schema Information. In *Spinning the Semantic Web*, S.197-S.222, Massachusetts Institute of Technology, 2003.
- [BrPa98]
Brin, Sergey; Page, Lawrence: The anatomy of a large-scale hyper-textual Web search engine. In the 7th International World Wide Web Conference, Brisbane, Australia, 1998. <http://www.stanford.edu/class/cs240/readings/google.pdf>, Abruf am 2005-08-08. <http://www-db.stanford.edu/~backrub/google.html>, Abruf am 2005-08-08.
- [BrSc85]
Brachman, R. J.; Schmolze, J. G.: An overview of the KL-ONE knowledge representation system. *Cognitive Science*, 9(2), pp. 171-216, 1985.
- [Burk00]
Burke, Robin: Semantic ratings and heuristic similarity for collaborative filtering. <http://www.igec.umbc.edu/kbem/final/burke.pdf>, 2000, Abruf am 2005-06-04.

C

- [CaCo03]
Cannataro, Mario; Comito, Carmela: The Grid: A Data Mining Ontology for Grid Programming, Proceedings of SemPGRID '03. Budapest 2003. <http://www.isi.edu/~stefan/SemPGRID/proceedings/proceedings.pdf>, S. 115-134. 2003-05-20, Abruf am 2005-06-03.
- [ChSh03]
Chen, Yi-Shin; Shahabi, Cyrus: Improving User Profiles for E-Commerce by Genetic Algorithms. <http://www.csee.umbc.edu/%7Ekolari1/Mining/papers/YodaGA.pdf>, 2003-10-22, Abruf am 2005-06-03.
- [CKGR05]
China Knowledge Grid Research (CKG): Knowledge Grid Center. <http://kg.ict.ac.cn/default.asp>, Abruf am 2005-08-11.
- [Clem03]
Clemens, Tobias: Mobile Marketing – Grundlagen, Rahmenbedingungen und Praxis des Dialogmarketings über das Mobiltelefon. Verlag Dr. Müller, Düsseldorf 2003.

[Cogn00]

Cognos GmbH: Cognos Guide - Der Business Intelligence Leitfaden, Cognos GmbH (Hrsg.), 2. Auflage, Frankfurt 2000.

[CoMS99]

Cooley, R.; Mobasher, B.; Srivastava, J.: Data preparation for mining world wide web browsing patterns. *Journal of Knowledge and Information Systems*, 1999.

[Cove00]

Cover Pages: Technology Reports – Simple HTML Ontology Extensions (SHOE). <http://xml.coverpages.org/shoe.html>, Abruf am 2006-02-11.

D

[DaLP03]

Dave, Kushal; Lawrence, Steve; Pennock, David. M.: Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews.

<http://www.kushaldave.com/p451-dave.pdf>, 2003, Abruf am 2005-06-04.

[DaOS03]

Daconta, Michael C.; Obrst, Leo J.; Smith, Kevin T.: The Semantic Web: a guide to the future of XML, web services, and knowledge management. Indianapolis, Ind, 2003.

[Davi03]

Davies, John (Hrsg.): Towards the Semantic Web: ontology-driven knowledge management. Chichester 2003.

[Dear02]

Dearlove, Des: Die Bill Gates Methode – Die 10 Erfolgsgeheimnisse des reichsten Unternehmers der Welt. Wirtschaftsverlag Carl Ueberreuter, Wien-Frankfurt. 3. Auflage, 2003.

[DMOZ05]

Directory Mozilla (DMOZ): Open Directory Project <http://www.dmoz.com/about.html>, Abruf am 2005-11-25.

[DoRS05]

Domingue, John; Roman, Dumitru; Stollberg, Michael : Web Service Modeling Ontology (WSMO) – An Ontology for Semantic Web Services. Position paper at the W3C Workshop on Frameworks for Semantics in Web Services, Innsbruck.

http://www.w3.org/2005/04/FSWS/Submissions/1/wsmo_position_paper.html, 2005-06-10, Abruf am 2005-07-11, 2005-08-05.

[DSKT02]

Desikan, Prasanna; Srivastava, Jaideep; Kumar, Vipin; Tan, Pang-Ning: Hyperlink Analysis Techniques & Applications. Army High Performance Computing Center Technical Report, 2002.

<http://www-users.cs.umn.edu/~desikan/publications/survey.pdf>, Abruf am 2005-08-08.

E

[Eber04]

Eberhart, Andreas: Ontology-based Infrastructure for Intelligent Applications. Dissertation an der Naturwissenschaftlich-Technischen Fakultät, Saarbrücken, 2004. <http://scidok.sulb.uni-saarland.de/volltexte/2004/260/pdf/EberhartProfDrWolfgangWahlster.pdf>, Abruf am 2006-04-11.

[Eich00]

Eichhorn, Peter: „Das Prinzip Wirtschaftlichkeit“. Betriebswirtschaftlicher Verlag Dr. Th. Gabler GmbH, Wiesbaden 2000.

[EiVa03]

Eirinaki, Magdalini; Vazirgiannis, Michalis: Web Mining for Web Personalization http://www.csee.umbc.edu/%7Ekolari1/Mining/papers/TOIT-webmining_survey.pdf, 2003-10-24, Abruf am 2005-06-03.

F

[FaSt02]

Faatz, Andreas; Steinmetz, Ralf: Ontology Enrichment with Texts from the WWW. Darmstadt University of Technology, Darmstadt, 2002. <http://km.aifb.uni-karlsruhe.de/ws/semwebmine2002/papers/full/faatz.pdf>, Abruf am 2006-02-13.

[FHLW03]

Fensel, Dieter; Hendler, James; Lieberman, Henry; Wahlster, Wolfgang: Spinning the Semantic Web – Bringing the World Wide Web to Its Full Potential. Massachusetts Institute of Technology, 2003.

[FHLW05]

Fensel, Dieter (Hrsg.); Hendler, James; Lieberman, Henry; Wahlster, Wolfgang: Spinning the Semantic Web – Bringing the World Wide Web to Its Full Potential. MIT Press, Cambridge, Mass 2005.

[FiPY01]

Fix, Ulla; Poethe, Hannelore; Yos, Gabriele: Textlinguistik und Stilistik für Einsteiger, Leipziger Skripten, Einführungs- und Übungsbücher, Peter Lang GmbH, Europäischer Verlag der Wissenschaften Frankfurt am Main 2001.

[FiMi75]

Fleischer, Wolfgang; Michel, Georg: Stilistik der deutschen Gegenwartssprache. Leipzig 1975.

[FoHo99]

Fox, D.; Horster, P.: Datenschutz und Datensicherheit - Inhaltsfilterung und Jugendschutz im Internet. Wiesbaden 1999. <http://www.secorvo.de/publikationen/juschutz.pdf>, Abruf am 2005-06-03.

[FrPL03]

Frye, Christopher; Plusch, Mike; Liebermann, Henry: Static and Dynamic Semantics of the Web. In Spinning the Semantic Web, Massachusetts Institute of Technology, 2003.

[Fürn03]

Fürnkranz, Johannes: Web Structure Mining Exploiting the Graph Structure of the World-Wide Web. <http://www.csee.umbc.edu/%7Ekolari1/Mining/papers/oefai-tr-2002-33.pdf>, 2003-10-24, Abruf am 2005-06-03.

G

[GaSc02]

Gaul, Wolfgang; Schmidt-Thieme, Lars: Web Controlling und Recommendersysteme. In Handbuch Web Mining im Marketing - Konzepte, Systeme, Fallstudien. S.234-265.

- [Gate05]
Gates, Bill: Executive E-mail. <http://www.microsoft.com/mscorp/execmail/2005/02-03interoperability-print.asp>, 2005-02-03, Abruf am 2005-06-20.
- [GeCl02]
Gentsch, Peter; Claus, Stefan: Web Mining für die Personalisierung von e-Portalen. In Handbuch Web Mining im Marketing - Konzepte, Systeme, Fallstudien. S.420-436.
- [Gent02]
Gentsch, Peter: Personalisierung der Kundenbeziehung im Internet-Methoden und Technologien. In Handbuch Web Mining im Marketing - Konzepte, Systeme, Fallstudien. S.266-307.
- [Gero04]
Geroimenko, Vladimir: Dictionary of XML Technologies and the Semantic Web. Springer, London 2004.
- [Goog05]
Google Inc.: Webseite. <http://www.google.com>, Abruf am 2005-08-08.
- [Grub93]
Gruber, Tom R.: Toward principles for the design of ontologies used for knowledge sharing. Padua workshop on Formal Ontology, March 1993. <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>, Abruf am 2006-02-09.
- [Guha01]
Guha, R.: Open Rating System. http://www.w3.org/2001/sw/Europe/events/foaf-galway/papers/fp/open_rating_systems/wot.pdf, Abruf am 2005-07-11, 2005-08-05.
- [GVSR01]
Gentsch, Peter; Veth, Christian; Schinzer, Heiko D.; Roth, Michael; Mandzak, Peter; Bange, Carsten: Web-Personalisierung und Web-Mining für eCRM – 12 Software-Lösungen im Vergleich. Eine Studie des Business Application Research Centers, Oxygon Verlag GmbH, 2001.
-
- ## H
- [HaCh03]
Han, Jiawei; Chang, Chen-Chuan Kevin: Data Mining for Web Intelligence. <http://www.csee.umbc.edu/%7Ekolari1/Mining/papers/computer02.pdf>, 2003-10-24, Abruf am 2005-06-03.
- [HaSZ02]
Haasis, Klaus; Strommer, Walter; Zerfuß, Ansgar (Hrsg.): Digitale Wertschöpfung – Internet und E-Business als Chance für den Mittelstand. Heidelberg : dpunkt – Verlag, 2002.
- [HeHe00]
Heflin, Jeff; Hender, James: Semantic Interoperability on the Web. <http://www.cs.umd.edu/projects/plus/SHOE/pubs/extreme2000.pdf>, 2000, Abruf am 2005-06-05.
- [HiMW02]
Hippner, Hajo; Merzenich, Melanie; Wilde, Klaus: Handbuch Web Mining im Marketing – Konzepte, Systeme, Fallstudien. Verlag Vieweg, 1. Auflage, September 2002.
- [HKMP03]
Henriques, Pedro Rangel; Kosar, Tomaz; Mernik, Marjan; Pereira, Maria Joao Varanda; Zumer, Viljem: Grammatical Approach to Problem Solving. SRCE University Computing Centre, University of Zagreb 2003.
-

http://wiki.di.uminho.pt/wiki/pub/Voda/WebHome/ITI-grammatical_approach.pdf, Abruf am 2005-08-11.

[HKMW01]

Hippner, Hajo; Küsters, Ulrich; Meyer Matthias; Wilde, Klaus: Handbuch Data Mining im Marketing – Knowledge Discovery in Marketing Databases. Braunschweig, Vieweg 2001.

[HoWe02]

Hofmann, Hanno; Weingärtner, Stefan: Aufbau einer e-Intelligence-Architektur für das Personality-Portal koepfe.de. In Handbuch Web Mining im Marketing - Konzepte, Systeme, Fallstudien. S.438-452.

I

[Illi98]

Illik, J.A: Electronic Commerce – eine systematische Bestandsaufnahme. HMD, 35.Jg. 1998, Heft 199, S.10-24

[ISO05]

ISO : International Organization for Standardization
<http://www.iso.org/iso/en/ISOOnline.frontpage>, Abruf am 2005-10-30.

J

[JiMo03]

Jin, Xin; Mobasher, Bamshad: Using Semantic Similarity to Enhance Item-based Collaborative Filtering. <http://www.csee.umbc.edu/%7Eekolari1/Mining/papers/JM03.pdf>, 2003-10-22, Abruf am 2005-06-03.

K

[Kahm01]

Kahmann, M. (Hrsg): Report Mobile Business – Neue Wege zum mobilen Kunden, Düsseldorf 2001.

[KJNY03]

Krishnapuram, Raghu; Joshi, Anupam; Nasraoui, Olfa; Yi, Liyu: Low-Complexity Fuzzy Relational Clustering Algorithms for Web Mining.
<http://www.csee.umbc.edu/%7Eekolari1/Mining/papers/krishnapuram01lowcomplexity.pdf>, 2003-10-14, Abruf am 2005-06-03.

[Klus01]

Klusch Matthias: Intelligente Informationsagenten für Wissensentdeckung und Data Mining im Internet. Rest siehe [HKMW01]

- [KoBl03]
Kosala, Raymond; Blockeel, Hendrik: Web Mining Research: A Survey
<http://www.csee.umbc.edu/%7Ekolari1/Mining/papers/kosala.pdf>, 2003-10-24, Abruf am 2005-06-03.
- [Koch03]
Koch, Michael: Community-Unterstützungssysteme – Architektur und Interoperabilität. Technische Universität München, 2003. <http://www.communixx.de/files/Koch2003d.pdf>, Abruf am 2006-04-10.
- [KoPa03]
Kohavi, Ron; Parekh, Rajesh: Ten Supplementary Analyses to Improve E-commerce Web Sites.
<http://www.csee.umbc.edu/%7Ekolari1/Mining/papers/supplementaryAnalyses.pdf>, 2003-10-22, Abruf am 2005-06-03.
- [KrBh04]
Krishnamoorthy, Srikumar; Bharat, Bhasker: Applications of recommender systems in target selection.
<http://proquest.umi.com/pqdweb?index=1&did=735681801&SrchMode=1&sid=2&Fmt=6&VInst=PROD&VType=PQD&RQT=309&VName=PQD&TS=1119963505&clientId=37603>, 2004-05-25. Abruf am 2005-06-28. (ProQuest)
- [KrSo98]
Krause, Jörg; Somm, Felix: Online-Marketing – Die perfekte Strategie für Ihren Internet-Auftritt. Carl Hanser Verlag München Wien, 1998.
- [KrSr04]
Krishnamoorthy, Srikumar: A Framework of Agent-based Personalized Recommender System for E-Commerce
<http://proquest.umi.com/pqdweb?index=2&did=816657351&SrchMode=1&sid=2&Fmt=6&VInst=PROD&VType=PQD&RQT=309&VName=PQD&TS=1119963505&clientId=37603>, Abruf am 2005-06-28. (ProQuest)
- L**
- [Lang94]
Langenohl, T.: Systemarchitekturen elektronischer Märkte, Diss., St. Gallen 1994.
- [LeGu90]
Lenat, D. B.; Guha, R. V.: Building Large Knowledge-Based Systems: Representation and Inference in the CYC Project. Addison-Wesley, Reading, Massachusetts, 1990.
- [LiBe01]
Linoff, Gordon, S.; Berry, Michael J.A.: Mining the Web – transforming Customer Data into Customer Value. 2001.
- [Lima05]
Liman Literatur-Manager: die bewährte Literaturverwaltungssoftware
<http://www.liman.de/>, Abruf am 2005-11-18.
- [LiSc00]
Linton, Frank; Schäfer, Hans-Peter: Recommender Systems for Learning: Building User and Expert Models through Long-Term Observation of Application Use.
<http://proquest.umi.com/pqdweb?index=10&did=403857531&SrchMode=1&sid=2&Fmt=10&VInst=PROD&VType=PQD&RQT=309&VName=PQD&TS=1119963505&clientId=37603>, Abruf am 2005-06-28.
-

M

- [MaBr92]
MacGregor, R.; Brill, D.: Recognition algorithms for the LOOM classifier. In Proc. AAAI-92. The MIT Press. 774—779, 1992.
- [Maes94]
Maes, P.: Agents that reduce work and information overload. Communications of the ACM, 1994
- [MaYB89]
Malone, T.W.; Yates, J.; Benjamin, R.I.: The Logic of Markets. Harvard Business Review, 67. Vol. 1989, S. 166 – 170.
- [McGu02]
McGuinness, L. Deborah: Ontologies Come of Age. In Dieter Fensel, J im Hendler, Henry Lieberman, and Wolfgang Wahlster, editors. Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential. MIT Press, 2002.
[http://www.ksl.stanford.edu/people/dlm/papers/ontologies-come-of-age-mit-press-\(with-citation\).htm](http://www.ksl.stanford.edu/people/dlm/papers/ontologies-come-of-age-mit-press-(with-citation).htm), Abruf am 2005-11-18.
- [McGu03]
McGuinness, Deborah L.: Ontologies Come of Age. In Spinning the Semantic Web, S.171-S.195, Massachusetts Institute of Technology, 2003.
- [MCKR04]
Mernik, Marjan; Crepinsek, Matej; Kosar, Tomaz; Rebernak, Damijan; Zumer, Viljem: Grammar-Based Systems: Definition and Examples.
<http://marcel.uni-mb.si/marjan/Informatika04.pdf>, 2004-01-30, Abruf am 2005-06-03.
- [McPi04a]
McGuinness, L. Deborah; Pinheiro da Silva, Paulo: Explaining Answers from the Semantic Web: The Inference Web Approach. Journal of Web Semantics. Vol.1 No.4, S. 397-413, Oktober 2004.
http://www.ksl.stanford.edu/people/pp/papers/McGuinness_KSL_04_03.pdf, Abruf am 2005-08-09.
- [McPi04b]
McGuinness, L. Deborah; Pinheiro da Silva, Paulo: Trusting Answers on the Web. In Mark T. Maybury, editor, New Directions in Question Answering. Chapter 21, AAAI/MIT Press, October 2004.
http://www.ksl.stanford.edu/people/pp/papers/McGuinness_QA_2003.pdf, Abruf am 2005-08-09.
- [McRI95]
McGuinness, D.L; Resnick A. L.; Isbell, C.: Description Logic in practice: A CLASSIC application. In Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence. International Joint Committee on Artificial Intelligence, August 1995.
- [MiÖs05]
Microsoft Österreich: Fakten zu Windows und Linux. <http://www1.microsoft.at/fakten/>, Abruf am 2005-06-20.
- [MoAu05a]
Mobilkom Austria AG und Co KG: Webseite. <http://www.a1.net/>, 2005, Abruf am 2005-12-01.
- [MoAu05b]
Mobilkom Austria AG & Co KG: Das Unternehmen. <http://www.mobilkomaustria.com/>, Abruf am 2005-12-27.

- [mobi05]
mobileup!.de: Das Online Magazin für Handy und Mobilfunk. <http://www.mobileup.de/handy-lexikon/b.html>, Abruf am 2005-11-17.
- [MoCS00]
Mobasher, B.; Cooley, R.; Srivastava, J.: Automatic personalization based on web usage mining. *Commun. ACM*, 43, August 2000, S. 142-151.
- [MoEN00]
Moh, Chuang-Hue; Lim, Ee-Peng; Ng Wee-Keong: DTD-Miner: A Tool for Mining DTD from XML Documents. *WECWIS 2000*: 144-151. <http://pmg.lcs.mit.edu/~chmoh/pubs/wecwis.pdf>, Abruf am 2005-08-08.
- [MoJZ04]
Mobasher, Bamshad; Jin, Xin; Zhou, Yanzan: Semantically Enhanced Collaborative Filtering on the Web. <http://maya.cs.depaul.edu/~mobasher/papers/ewmf04-web/>, 2004, Abruf am 2005-06-04.
- [MuAB00]
Mulvenna, M. D.; Anand, S. S.; Buchner, A. G.: Personalization on the net using web mining. *Commun. ACM*, 43, August 2000, S.123-125.
- [MuSt05]
MusicStrands, Inc: Recommendation and personalization technology for music business. <http://indy.musicstrands.com/corp/products/recommender.html>, Abruf am 2005-12-14.

N

- [NoKl04]
Noy, F. Natalya; Klein, Michel: Ontology Evolution: Not the Same as Schema Evolution. In *Knowledge and Information Systems*, 6(4):428-440. 2004
<http://www.cs.vu.nl/~mcaklein/papers/NoyKlein.pdf>, Abruf am 2005-08-11.

O

- [OneG05]
One GmbH: Unternehmensprofil. www.one.at, Abruf am 2005-12-27.
- [Open05]
Opensource.org: The Open Source Definition – Version 1.9.
<http://www.opensource.org/docs/definition.php>, 2005, Abruf am 2005-06-11.

P

- [PiMc04]
Pinheiro da Silva, Paulo; McGuinness, L. Deborah: A Proof Markup Language for Semantic Web Services. Stanford University, 2004.
ftp://ftp.ksl.stanford.edu/pub/KSL_Reports/KSL-04-01.pdf, Abruf am 2005-08-09.

[PiMM03]

Pinheiro da Silva, Paulo; McGuinness, L. Deborah; McCool, Rob: Knowledge Provenance Infrastructure. IEEE Data Engineering Bulletin Vol.26 No.4, S. 26-32, Dezember 2003.
http://www.ksl.stanford.edu/people/pp/papers/PinheirodaSilva_DEBULL_2003.pdf, Abruf am 2005-08-09.

[Plu98]

Plu, M.: Software technologies for building agent based systems in telecommunication networks. In: N.R. Jennings and M. Wooldridge: Agent Technology, Springer, 1998.

[PRSB94]

Preece, J. Jennifer; Rogers, Yvonne; Sharp, H.; Benyon, David; Holland, Simon; Carey, T.: Human-Computer Interaction. Addison Wesley, 1994.

R

[Rahm05]

Rahm, Erhard: Web Usage Mining. <http://mordor.prakinf.tu-ilmenau.de/papers/dbspektrum/dbs-02-75.pdf>, Abruf am 2005.-08-08.

[Refs05]

Refsnes Data: W3 Schools. <http://www.w3schools.com/>, Abruf am 2005-12-09.

[ReMö97]

Reichwald, Ralf; Möslein, Kathrin: Chancen und Herausforderungen für neue unternehmerische Strukturen und Handlungsspielräume in der Informationsgesellschaft. In Picot, Arnolde (Hrsg.): Telekooperation und virtuelle Unternehmen. R.v. Decker's Verlag, Heidelberg 1997.

[ReWe06]

Reimon, Michel; Weixler, Helmut: Die sieben Todsünden der EU. Verlag Carl Ueberreuter, Wien, 2006.

[Robo05]

Robotstxt.org: <http://www.robotstxt.org>, Abruf am 2005-08-08.

[RoVo02]

Roth, Michael; Voss, Jan-Martin: Web Mining Application Services Providing – Erfahrungen und Erfolgsfaktoren. In Handbuch Web Mining im Marketing - Konzepte, Systeme, Fallstudien. S.472-485.

S

[Säub02]

Säuberlich, Frank: Der Web Mining Prozess. In Handbuch Web Mining im Marketing - Konzepte, Systeme, Fallstudien. S.105-123.

[ScBM99]

Schinzler, H.; Bange, C.; Mertens, H.: Data Warehouse und Data Mining. 2. Auflage, München 1999.

- [SCDT00]
Srivastava, Jaideep; Cooley, Robert; Deshpande, Mukund; Tan, Pang-Ning: Web Usage Mining - Discovery and Applications of usage patterns from Web Data. SIGKDD Explorations, Vol 1, Issue 2, 2000. <http://terral.lsi.uned.es/WebMining/Tema5.Usos/srivastava2000.pdf>, Abruf am 2005-08-08.
- [SCDT03]
Srivastava, Jaideep; Cooley, Robert; Deshpande, Mukund; Tan, Pang-Ning: Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data. <http://www.csee.umbc.edu/%7Eekolari1/Mining/papers/Srivastava2000.pdf>, 2003-10-22, Abruf am 2005-06-03.
- [Scha03]
Schackmann, Jürgen: Ökonomisch vorteilhafte Individualisierung und Personalisierung – Eine Analyse unter besonderer Berücksichtigung der Informationstechnologie und des Electronic Commerce. Verlag Dr. Kovac 2003.
- [ScJu01]
Schmich, P.; Juszczak, L.: Mobile Marketing Verlust der Privatsphäre oder Gewinn für den Verbraucher. In: Report Mobile Business – Neue Wege zum mobilen Kunden (S.81), Düsseldorf 2001.
- [Seco00]
Secorvo Security Consulting GmbH: Studie „Jugendschutz und Filtertechnologien im Internet“ <http://www.secorvo.de/projekt/jugendschutz.htm>, 2000, Abruf am 2005-06-03.
- [SeEd03]
Semantic Edge: Opportunities. http://www.semanticedge.com/opportunities/en_opport_know.htm, 2003, Abruf am 2005-11-17.
- [SGPM05]
Shvaiko, Pavel; Giunchiglia, Fausto; Pinheiro da Silva, Paulo; McGuinness, L. Deborah: Web Explanations for Semantic Heterogeneity Discovery. In Proceedings of the 2nd European Semantic Web Conference (ESWC 2005), Heraklion, Greece, May 29 to June 1, 2005. Springer 2005. ftp://ftp.ksl.stanford.edu/pub/KSL_Reports/KSL-04-02.pdf, Abruf am 2005-08-09.
- [ShBa03]
Shahabi, Cyrus; Banaei-Kashani, Farnoush: A Framework for Efficient and Anonymous Web Usage Mining Based on Client-Side Tracking. <http://www.csee.umbc.edu/%7Eekolari1/Mining/papers/Shahabi-WebKDD2001-BookChapter.pdf>, 2003-10-22, Abruf am 2005-06-03.
- [ShCh03]
Shahabi, Cyrus; Chen, Yi-Shin: Web Information Personalization: Challenges and Approaches. <http://www.csee.umbc.edu/%7Eekolari1/Mining/papers/DNIS2003.pdf>, 2003-10-22, Abruf am 2005-06-03.
- [ShYM05]
Sheizaf, Rafaeli; Yuval, Dan-Gur; Miri, Barak: Social Recommender Systems: Recommendations in Support of E-learning. <http://proquest.umi.com/pqdweb?index=0&did=806808621&SrchMode=1&sid=2&Fmt=6&VInst=PROD&VType=PQD&RQT=309&VName=PQD&TS=1119963505&clientId=37603>, Abruf am 2005-06-28. (ProQuest)

- [SJLX04]
Su, Z.; Jiang, J.; Liu, T.; Xie, G.T.; Pan, Y.: Market Intelligence: An entity-based system for managing market intelligence.
<http://proquest.umi.com/pqdweb?index=1&did=708283121&SrchMode=1&sid=2&Fmt=6&VInst=PROD&VType=PQD&RQT=309&VName=PQD&TS=1119973988&clientId=37603>, Abruf am 2005-06-28.
- [soko05]
software-kompetenz.de: Software-Engineering-Wissensdatenbank: Collaborative Filtering. Fraunhofer IESE. <http://www.software-kompetenz.de/?15854>, Abruf am 2005-12-14.
- [SoLe01]
Soto, Miche; Le Grand, Benedicte: XML Topic Maps and Semantic Web Mining.
<http://semwebmine2001.aifb.uni-karlsruhe.de/online/semwebmine08.pdf>, 2001, Abruf 2001-06-03.
- [SpGB03]
Spiekermann, Sarah; Grossklags, Jens; Berendt, Bettina: E-privacy in 2nd Generation E-Commerce: Privacy Preferences versus actual Behavior.
<http://www.csee.umbc.edu/%7Ekolari1/Mining/papers/spiekermann01eprivacy.pdf>, 2003-10-22. Abruf am 2005-06-03.
- [Spil01]
Spiliopoulou, Myra: Web Usage Mining: Data Mining über die Nutzung des Web. In Handbuch Data Mining im Marketing – Knowledge Discovery in Marketing Databases Braunschweig, Vieweg 2001.
- [Spil03]
Spiliopoulou, Myra: Web Usage Mining for Web Site Evaluation – Making a site better fit its users. <http://www.csee.umbc.edu/%7Ekolari1/Mining/papers/spiliopoulou.pdf>, 2003-10-22, Abruf am 2005-06-03.
- [SrDK03]
Srivastava, Jaideep; Desikan, Prasanna; Kumar, Vipin: Web Mining – Accomplishments & Future Directions.
<http://www.csee.umbc.edu/%7Ekolari1/Mining/papers/srivastava.pdf>, 2003-10-15, Abruf am 2005-06-03.
- [Steu98]
Steuck, Joachim W.: Geschäftserfolg im Internet – Das professionelle 1x1. Cornelsen Verlag, Berlin 1998.
- [Stan05]
Stanford Medical Informatics: Protégé – an open-source platform. <http://protege.stanford.edu/>, 2005. Abruf am 2005-12-06.
- [StHB03]
Stumme, Gerd; Hotho, Andreas; Berendt, Bettina: Usage Mining for and on the Semantic Web.
<http://www.csee.umbc.edu/%7Ekolari1/Mining/papers/stumme.pdf>, 2003-10-15, Abruf am 2005-06-03.

T

- [TaKu03]
Tan, Pang-Ning; Kumar, Vipin: Mining Association Patterns in Web Usage Data.
<http://www.csee.umbc.edu/%7Ekolari1/Mining/papers/184.pdf>, 2003-10-22, Abruf am 2005-06-03.

[TeHV01]

Terzi, Evimaria; Hacid, Mohand-Said; Vakali, Athena: Modeling and Querying Web Data: A Constraint-Based Logic Approach. <http://www.csee.umbc.edu/%7Ekolari1/Mining/papers/vakali-chapter.pdf>, 2001-10-18, Abruf am 2005-06-03.

[Torr00]

Torrent: Driving e-Commerce Profitability From Online and Offline Data – A Torrent Systems White Paper.

<http://www.csee.umbc.edu/%7Ekolari1/Mining/papers/torrent.pdf>, 2000, Abruf am 2005-06-03.

[TuLi02]

Turney, P. D.; Littman, M. L.: Unsupervised Learning of Semantic Orientation from a Hundred-Billion-Word Corpus. Technical Report ERB-1094, National Research Council Canada, Institute for Information Technology, 2002. <http://cogprints.org/2322/00/ERB-1094.pdf>, Abruf am 2005-08-11.

U

[Unic06]

Unicode Inc.: Unicode Home Page. <http://www.unicode.org>, Abruf am 2006-02-04.

[UNSP03]

United Nations Standard Products and Services Code (UNSPSC). <http://www.unspsc.org>, Abruf am 2005-11-18.

W

[Walt01]

Walther, Ralf: Web Mining, Informatik Spektrum – Gesellschaft für Informatik e.V.

<http://www.mindup.de/html/web-mining.html>, 2001, Abruf am 2005-06-03.

[WaMa05]

Walter MasterManagement GmbH: Unternehmensprofil. www.mastermanagement.at. Abruf am 2005-10-10.

[WeVe98]

Weihmayer, R; Velthuijsen, H.: Intelligent agents in telecommunications. In: N.R. Jennings and M. Wooldridge: Agent Technology, Springer, 1998.

[Wiel04]

Wielinga B. J: Vorfremde auf die Entwicklung von semantischen Webdiensten. Department of Social Science Informatics (SWI), University of Amsterdam. <http://www.innovations-report.de/html/berichte/informationstechnologie/bericht-29562.html>, Abruf am 2006-02-09.

[Wiki05]

Wikipedia: die freie Enzyklopädie. <http://de.wikipedia.org/> 2005, Abruf am 2005-06-03.

[WiMü02]

Widhalm, Richard; Mück, Thomas: Topic Maps – semantische Suche im Internet. Springer, Berlin 2002.

[WMPM05]

Welty, Christopher; Murdock, J. William; Pinheiro da Silva, Paulo; McGuinness, L. Deborah: Tracking Information Extraction from Intelligent Documents. In Proceedings of the 2005 International Conference on Intelligence Analysis (IA 2005), McLean, VA, USA, 2-6 May, 2005. http://iw.stanford.edu/documents_papers.html, Abruf am 2005-08-09.

[W3C99]

W3C Recommendation: Resource Description Framework (RDF) Model and Syntax Specification. 1999-02-22. <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/>, Abruf am 2005-12-06.

Z

[ZaPM05]

Zaihrayeu, Ilya; Pinheiro da Silva, Paulo; McGuinness, L. Deborah: IWTrust: Improving User Trust in Answers from the Web. Proceedings of 3rd International Conference on Trust Management (iTrust2005), Springer, Rocquencourt, France, 2005. http://www.ksl.stanford.edu/people/pp/papers/Zaihrayeu_iTrust_2005.pdf, Abruf am 2005-08-09.

[ZaWH04]

Zan, Huang; Wingyan, Chung; Hsinchun, Chen: A Graph Model for E-Commerce Recommender Systems <http://proquest.umi.com/pqdweb?index=3&did=525180621&SrchMode=1&sid=2&Fmt=10&VInst=PROD&VType=PQD&RQT=309&VName=PQD&TS=1119963505&clientId=37603>, Abruf am 2005-06-28.

[Zieg05]

Ziegler, Cai-Nicolas: Towards Decentralized Recommender Systems. Dissertation, 2005-06-13. <http://www.informatik.uni-freiburg.de/~chiegler/papers/A4-Thesis.pdf>, Abruf am 2005-12-14.

[Zipk01]

Zipkin, Paul: The limits of mass customization. In: Sloan Management Review 42 (2001), S. 81 – 87.

ABKÜRZUNGSVERZEICHNIS

ANSI-Code	Code of American National Standards Institute
ANSI-Protokoll	Protocoll of American National Standards Institute
API	Application Programming Interface
ASCII	American Standard Code for Information Interchange
ASP	Application Service Providing
BSC	Base Station Controller (Vermittlungsstelle)
BTS	Base Transceiver Station (Mobilfunkanlage, Basisstation)
BWL	Betriebswirtschaftslehre
B2C	Business-to-Consumer
CCC	Call and Communication Center
CF	Collaborative Filtering
CRM	Customer Relationship Management
eCRM	Electronic CRM
DAML	Darpa Agent Markup Language
DAML-O	Darpa Agent Markup Language for Ontologies
DARPA	Defence Advanved Research Projects Agency
DCMI	Dublin Core Metadata Initiative
D&B	Dunn & Bradstreet
E-Commerce	Electronic Commerce
EDGE	Enhanced Data Rates for Global Evolution
EDIFACT	United Nations initiative Electronic Data Interchange for Administration, Commerce, and Transport
EMS	Enhanced Message Service
F&E	Forschung und Entwicklung
GPRS	General Packet Radio Service
GSM	Global System for Mobile Communications
HTML	HyperText Markup Language
IBROW	Intelligent Brokering Service for Knowledge-Component Reuse
KDD	Knowledge Discovery in Databases
LBS	Location Based Service
LDAP	Lightweight Directory Access Protocol
mCRM	Mobiles CRM
MMS	Multimedia Messaging Service

MSC	Mobile Switching Center (Vermittlungsstelle)
OIL	Ontology Inference Layer, Ontology Interchange Layer
OKBC	Open Knowledge Base Connectivity protocol
OWL	Web Ontology Language (OWL nicht WOL!)
PC	Personal Computer
PDA	Personal Digital Assistent
RDF	Resource Description Framework
RDFS	Resource Description Framework Schema
ROI	Return on Investment
SGML	Standard Generalized Markup Language
SHOE	Simple HTML Ontology Extension
SMS	Short Message Service
SW	Semantic Web
UMTS	Universal Mobile Telecommunications System
UNDP	United Nations Development Programme
UNSPSC	United Standard Products and Services Code
UPML	Unified Problem-solving Method-development Language
URI	Uniform Resource Identifiers
URL	Unified Resource Locator
VWL	Volkswirtschaftslehre
WAP	Wireless Application Protocol
WML	Wireless Markup Language
WSDL	Web Services Description Language
WWW	World Wide Web
W3C	World Wide Web Consortium
XML	eXtensible Markup Language
XHTML	eXtensible HyperText Markup Language
XTM	XML Topic Map

BILDVERZEICHNIS

Bild 1 Sprachenarchitektur nach Fensel et al. [FHLW03].....	19
Bild 2 Entwicklung der Web-Sprachen	21
Bild 3 Beispiel einer SHOE-Ontologie [FHWL03].....	27
Bild 4 Beispiel einer SHOE-Instanz [FHWL03]	28
Bild 5 Art des Zugriffs auf eine Webseite [Spil01].....	33
Bild 6 Beispiele für verschiedene Web-Zugriffe	35
Bild 7 Methoden des Web Mining [HKMW01].....	40
Bild 8 Methoden der individualisierten Empfehlungssysteme	46
Bild 9 Kapitel-Übersicht.....	52
Bild 10 Instrumente der Informationstechnologie auf drei Ebenen.....	54
Bild 11 Ziele der Einflussfaktoren.....	54
Bild 12 Pyramide der informationstechnologischen Instrumente.....	55
Bild 13 Gesprächsverbindung zwischen zwei Handy-Benutzern.....	58
Bild 14 Unternehmen als Schnittstelle zwischen Technologie und Benutzer.....	60
Bild 15 Customer Relationship Management im Web	63
Bild 16 Abstraktion eines Web-Vorgangs (Maschine).....	63
Bild 17 Abstraktion eines Web-Vorgangs (Daten).....	65
Bild 18 Abstraktion eines Web-Vorgangs (Mensch).....	67
Bild 19 Maßnahmen der Webseiten-Optimierung [Bens02]	70
Bild 20 Ursache-Wirkung Beziehung.....	74
Bild 21 Direkter und indirekter Informationspfad.....	76
Bild 22 Ziele und Instrumente des CRM	78

TABELLENVERZEICHNIS

Tabelle 1 Anbieter der Web-Dienste.....	14
Tabelle 2 Berners-Lee's Architektur des Semantic Web [McGu02]	17
Tabelle 3 Geschäftsprozessanalyse am Beispiel des Bestellprozesses [Bens02]	72

ANHANG

1 Die vollständige WSDL 1.2 Syntax

```
<wsdl:definitions name="nmtoken"? targetNamespace="uri">
  <import namespace="uri" location="uri"/>
  <wsdl:documentation .... />

  <wsdl:types>
    <wsdl:documentation .... />
    <xsd:schema .... />
  </wsdl:types>

  <wsdl:message name="ncname">
    <wsdl:documentation .... />
    <part name="ncname" element="qname"? type="qname"?/>
  </wsdl:message>

  <wsdl:portType name="ncname">
    <wsdl:documentation .... />

    <wsdl:operation name="ncname">
      <wsdl:documentation .... />

      <wsdl:input message="qname">
        <wsdl:documentation .... />
      </wsdl:input>

      <wsdl:output message="qname">
        <wsdl:documentation .... />
      </wsdl:output>

      <wsdl:fault name="ncname" message="qname">
        <wsdl:documentation .... />
      </wsdl:fault>

    </wsdl:operation>
  </wsdl:portType>

  <wsdl:serviceType name="ncname">
    <wsdl:portType name="qname"/>
  </wsdl:serviceType>

  <wsdl:binding name="ncname" type="qname">
    <wsdl:documentation .... />

    <wsdl:operation name="ncname">
      <wsdl:documentation .... />
    </wsdl:operation>
  </wsdl:binding>
</wsdl:definitions>
```

```
<wsdl:input>
  <wsdl:documentation .... />
</wsdl:input>

<wsdl:output>
  <wsdl:documentation .... />
</wsdl:output>

<wsdl:fault name="ncname">
  <wsdl:documentation .... />
</wsdl:fault>

</wsdl:operation>

</wsdl:binding>

<wsdl:service name="ncname" serviceType="qname">
  <wsdl:documentation .... />

  <wsdl:port name="ncname" binding="qname">
    <wsdl:documentation .... />
  </wsdl:port>

</wsdl:service>

</wsdl:definitions>
```

(Quelle: http://www.w3schools.com/wsdl/wsdl_syntax.asp)