



acatech DISKUTIERT

> WIE ARBEITEN DIE SUCHMASCHINEN VON MORGEN?

INFORMATIONSTECHNISCHE, POLITISCHE
UND ÖKONOMISCHE PERSPEKTIVEN

FRIEDEMANN MATTERN (Hrsg.)

Bibliografische Information der Deutschen Bibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.ddb.de> abrufbar.

ISSN 1861-9924/ISBN 978-3-8167-7526-3

Alle Rechte vorbehalten

Dieses Werk ist einschließlich aller seiner Teile urheberrechtlich geschützt. Jede Verwertung, die über die engen Grenzen des Urheberrechtsgesetzes hinausgeht, ist ohne schriftliche Zustimmung von acatech unzulässig und strafbar. Dies gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen sowie die Speicherung in elektronischen Systemen.

© acatech – Deutsche Akademie der Technikwissenschaften, 2008

Geschäftsstelle	Hauptstadtbüro
Residenz München	
Hofgartenstraße 2	Jägerstraße 22/23
80539 München	10117 Berlin
Telefon +49 (0) 89/5 20 30 90	Telefon +49 (0) 30/39 88 50 71
Telefax +49 (0) 89/5 20 30 99	Telefax +49 (0) 30/39 88 50 72

E-Mail: info@acatech.de
Internet: www.acatech.de

Redaktion: Dr. Holger Jens Schnell, Dr. Jens Pape

Umschlaggestaltung: klink, liedig werbeagentur gmbh

Satz/Layout: Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS, Sankt Augustin

Herstellung und Produktion: Fraunhofer IRB Verlag, Stuttgart

Printed in Germany

Verlag und Vertrieb:
Fraunhofer IRB Verlag
Fraunhofer-Informationszentrum Raum und Bau IRB
Nobelstraße 12
70569 Stuttgart

Postfach 80 04 69
70504 Stuttgart

Telefon +49 (0) 711/970 25 00
Telefax +49 (0) 711/970 25 08
E-Mail: irb@irb.fraunhofer.de
Internet: www.IRBbuch.de

acatech DISKUTIERT

> WIE ARBEITEN DIE SUCHMASCHINEN VON MORGEN?

INFORMATIONSTECHNISCHE, POLITISCHE
UND ÖKONOMISCHE PERSPEKTIVEN

FRIEDEMANN MATTERN (Hrsg.)

> INHALT

> Vorwort	9
> Suchmaschinen – Eine kurze Einführung Friedemann Mattern	11
1 Orientierungshilfen im Hypertextdschungel	11
2 Der Traum von der Antwortmaschine	12
3 Wirtschaftliche Bedeutung von Suchtechnologie	14
4 Macht und Verantwortung der Suchmaschinen	16
5 Die Suchmaschinen von morgen	19
6 Leben im Zeitalter des „Instant Search“	20
7 Literatur	21
> Suchmaschinenforschung: Überblick und Systematisierung eines interdisziplinären Forschungsfeldes Marcel Machill/Markus Beiler/Martin Zenker	23
1 Einleitung	23
2 Suchmaschinenpolitik und -regulierung	25
3 Suchmaschinenökonomie	29
4 Suchmaschinen und Journalismus	32
5 Technik und Qualität von Suchmaschinen	36
6 Nutzerverhalten und -kompetenz	40
7 Literatur	45
> Von Suchmaschinen zu Antwortmaschinen: Semantische Technologien und Benutzerpartizipation im Web 3.0 Wolfgang Wahlster	59
1 Einleitung	59
2 Von syntaktischen Suchmaschinen zu semantischen Antwortmaschinen	60
3 Web 3.0 = Semantisches Web + Web 2.0	63

4	SmartWeb: Eine multimodale und mobile Antwortmaschine	64
5	Multimodale und lokationssensitive Antwortmaschinen	68
6	Semantische Videosuche und Peer-to-Peer-Suchmaschinen	70
7	Konklusion	72
8	Danksagung	73
9	Literatur	73
>	Wirtschaftliche Aspekte der Suchtechnologie:	
	Es geht um (viel) mehr als (nur) Suche	
	Stefan Wess/Jasmin Franz	77
1	Warum suchen wir?	77
2	Wie suchen wir heute?	78
3	Wie wir finden	81
4	Wie funktioniert das?	82
5	Praxisbeispiel „Knox“: Vertikale Suche und Enterprise 2.0 in der betrieblichen Praxis	84
6	Fazit	86
7	Literatur	87
>	Die Bedeutung von Suche im Online-Geschäft	
	Michael Buschmann	89
1	Einleitung	89
2	Bedeutung der Suche für den Internet-Nutzer	89
3	Bedeutung der Suche für Online-Unternehmen	90
4	Betrieb eines eigenen Suchportals unter „suche.t-online.de“	91
5	Einbindung von Sponsored Links im Kontext von redaktionellen Artikeln	94
6	Search Engine Marketing (SEM)	94
7	Search Engine Optimization (SEO)	96
8	Enabling-Suche	98
9	Fazit	99
10	Literatur	100

> Volltextsuchdienste – was darf/soll wie hinein und hinaus?	
Einige Anmerkungen zum Opium der Informationsgesellschaft	
Rainer Kuhlen	101
1 Zusammenfassung	101
2 In Richtung Dismediation und Retailgeschäft	101
3 Im Kontext anderer Volltextnachweis- oder Lieferdienste	104
3.1 AutoDoc – Mehrwerteffekte mit hohen Kosten	105
3.2 Dokumentlieferung im Bibliotheksverbund	106
3.3 Neue Dienstleistungen auch in öffentlichen Bibliotheken – das Konzept „Onleihe“	107
3.4 Retaildienst von ScienceDirect und Scirus	110
3.5 CrossRef	112
4 Suchdienste für (retrodigitalisierte) Buchbestände	114
4.1 Google Book Search (GBS)	114
4.2 GBS Google Print	115
4.3 Volltextsuche-Online (VTO/libreka)	116
4.4 GBS Library Program	117
5 Einige Trends	119
6 Literatur	124
> Auf dem Weg in die Google-Gesellschaft?	
Rudi Schmiede	127
1 Einleitung	127
2 Worum es geht: der Umgang mit dem Wissen der Welt	127
3 Informationeller Kapitalismus und Google	128
4 Informatisierung und Wissen	129
5 Kompetenzen, Chancen und Gefahren	130
6 Fazit	132
7 Literatur	132
> Über die Autoren	135

> VORWORT

Das Internet ist innerhalb weniger Jahre zu einer selbstverständlichen und nahezu unverzichtbaren Infrastruktur unserer Gesellschaft geworden – im geschäftlichen wie auch im privaten Umfeld. Den wichtigsten Dienst des Internets stellen dabei neben E-Mail die Suchmaschinen dar, die als Universalschnittstelle mit der digitalen Welt agieren und dem Nutzer als Antwort auf seine Stichwort-Anfrage in Sekundenbruchteilen eine Linkliste mit passenden Dokumenten im Web liefern. Bald aber werden nicht nur immer mehr, sondern auch ganz unterschiedliche Daten und Informationen den Weg ins Internet finden: Neben Fotos und Videos auch eingescannte Bücher, medizinische Datenbanken sowie viele von Sensoren in Realzeit generierte Daten zum Beispiel über die Umwelt, den Zustand von Produkten, den momentanen Ort von Gegenständen und Personen oder sogar über die eigene Gesundheit. Die Suche in solchen semantisch komplexen Datenbereichen erfordert „intelligente“ Mechanismen, die weit über heutige stichwortbasierte Verfahren hinausgehen, damit aus den Daten nützliche Informationen und letztlich Wissen und Wert entsteht.

Wie aber sehen die Suchmaschinen der nächsten Generation aus, und wo muss dazu im Bereich der Informationstechnologie weiter geforscht werden, damit aus den Suchmaschinen letztlich „Antwortmaschinen“ werden? Was bedeutet generell die Suchtechnologie für die Wirtschaft und welche unternehmerischen Chancen bieten sich? Welche gesellschaftlichen Herausforderungen – etwa hinsichtlich Medienkompetenz, einseitiger Information oder des Schutzes der mitprotokollierten Nutzerdaten – bergen Suchmaschinen und welche Probleme ergeben sich aus der Marktdominanz einiger weniger Anbieter?

Mit diesen und weiteren Fragen beschäftigte sich das acatech Symposium „Mehr als Stichworte: Wie arbeiten die Suchmaschinen von morgen? Informationstechnische, politische und ökonomische Perspektiven“, das am 31. Mai 2007 in Berlin unter Beteiligung namhafter Experten aus Informatik und Informationstechnik, Medienwissenschaft sowie

Wirtschaft und Politik stattfand. Der vorliegende Tagungsband enthält die schriftlichen Ausarbeitungen der Vorträge und liefert damit einen aktuellen Querschnitt zu diesem wichtigen Thema, das aufgrund seiner Bedeutung und Brisanz in nächster Zeit sicherlich noch intensiver – auch unter Beteiligung von acatech – diskutiert werden dürfte.

Zürich, Dezember 2007

Friedemann Mattern

Leiter des acatech Themennetzwerks Informations- und Kommunikationstechnologie

> SUCHMASCHINEN – EINE KURZE EINFÜHRUNG

FRIEDEMANN MATTERN

1 ORIENTIERUNGSHILFEN IM HYPERTEXTDSCHUNGEL

Das Internet ist innerhalb weniger Jahre zu einer selbstverständlichen und nahezu unverzichtbaren Infrastruktur unserer Gesellschaft geworden – im geschäftlichen wie auch im privaten Umfeld. Dabei ist eine paradoxe Situation entstanden: Wir dürsten einerseits nach immer mehr Wissen und ertrinken andererseits doch in einem Meer von Information. Denn nicht nur die reine Datenmenge und die im Internet zugreifbaren Dokumente vermehren sich rasant, auch die Zahl der interessanten Websites, Blogs und Internet-Portale nimmt so rasch zu, dass man sich ohne Instrumente praktisch nicht mehr zurechtfinden kann. Als unentbehrliche Orientierungshilfen in diesem Hypertextdschungel dienen die Suchmaschinen, die uns in Sekundenbruchteilen als Resultat auf eine kurze Stichwort-Anfrage hin eine Liste mit Links zu passenden Dokumenten und Webseiten liefern, die nur noch angeklickt werden müssen. Suchmaschinen stellen daher mittlerweile neben E-Mail den wichtigsten Dienst des Internets dar. Sie bilden zusammen mit den Webbrowsern die Universalschnittstelle zur digitalen Welt.

Dennoch sind Suchmaschinen eine vergleichsweise junge Technologie. Sie entstanden erst Mitte der 1990er Jahre – praktisch zeitgleich mit den einfach zu bedienenden Browsern von Netscape und Microsoft, die das World Wide Web populär gemacht haben; simple Dokumentenkataloge, wie die zu Beginn der 1990er Jahre entstandenen und heute kaum noch bekannten ftp-Verzeichnisse „Archie“ oder „Veronica“ konnten das steigende Bedürfnis nach Orientierung im rasch wachsenden Netz nicht mehr erfüllen. Doch mit Ausnahme von Yahoo spielen heute die Suchmaschinen-Pioniere der „Prä-Google-Ära“ wie Excite, Lycos, AltaVista oder HotBot kaum noch eine Rolle; der kometenhafte Aufstieg der Firma Google vom 1998 gegründeten Start-up-Unternehmen hin zum Internet-Giganten innerhalb nur weniger Jahre ist mittlerweile schon legendär: Bereits im Jahr 2005 überholte Google mit einem Börsenwert von seinerzeit rund 80 Milliarden Dollar das bis dahin weltgrößte Medienunternehmen Time Warner.

Abb. 1: Suchmaschinen werden ubiquitär



Inzwischen stellen Suchmaschinen auch ein selbstverständlich genutztes „professionelles“ Rechercheinstrument dar, das von Journalisten, Schülern, Studenten, Wissenschaftlern, Marketingexperten, Personalfachleuten und Geheimdiensten gleichermaßen gern verwendet wird. Die Aussage eines Studenten, dass er ohne Google nicht gewusst hätte, wie er die Seminararbeit hätte schaffen sollen, ist durchaus glaubhaft! Wenn es aber ohne Google oder andere Suchmaschinen nicht mehr „zu gehen“ scheint, dann gewinnen Suchmaschinen offenbar eine weit über das reine Internet hinausgehende Bedeutung und werden zu einer Schlüsseltechnologie in der „realen Welt“. Gleichzeitig gewinnen Fragen danach, welche Informationen eine Suchmaschine eigentlich liefert, liefern soll oder liefern darf, wie die Qualität der Ergebnisse sichergestellt werden kann, wie das „Ranking“ der Suchergebnisse erfolgt etc., eine ökonomische und zunehmend auch eine kulturelle und politische Bedeutung.

2 DER TRAUM VON DER ANTWORTMASCHINE

Noch vor 15 Jahren wäre es sowohl in technischer als auch in ökonomischer Hinsicht kaum vorstellbar gewesen, dass Milliarden von weltweit verteilten Dokumenten indiziert werden und auf eine Anfrage hin jedem Nutzer im Bruchteil einer Sekunde eine passende Teilmenge davon kostenlos zur Verfügung gestellt wird. Kinder allerdings durften davon träumen. So hieß es in einem amerikanischen Kinderbuch¹ von 1964 beispielsweise über die fantastischen Möglichkeiten der Zukunft (vgl. Abb. 2): „You’re doing your homework. You’re stuck and you need some answers. So you get help from your answer machine. On a table next to you is part of the machine – a typewriter keyboard. When you push the correct keys, your questions will be answered on a screen on the wall.“ Und weiter: „You might ask (using the keyboard): WHO INVENTED THE PHONOGRAPH? And the machine would answer: THOMAS A. EDISON.“ Oder: „WHAT DID EDSION’S FIRST PHONOGRAPH

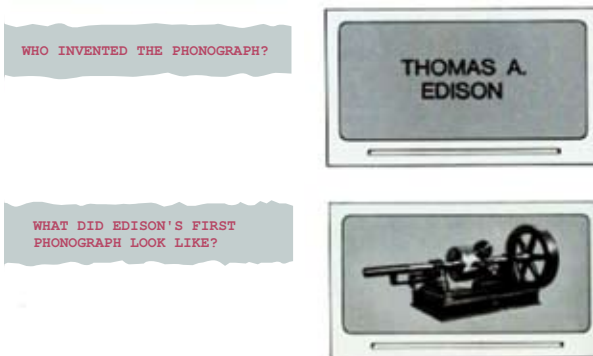
¹ Childcraft 1964.

LOOK LIKE?", und die „Antwortmaschine“ würde auf dem futuristischen Wanddisplay ein Bild des Phonographen zeigen (vgl. Abb. 3). „Someday, you might have an answer machine that could do all those things“, wird den Kindern 1964 versprochen.

Abb. 2: Die Antwortmaschine hilft bei den Hausaufgaben – ein Kindertraum von 1964



Abb. 3: Die Antwortmaschine in Aktion



Wie weit sind wir heute, gut 40 Jahre später, noch von der Realisierung einer solchen „Antwortmaschine“ entfernt? Wenn man sich das Ergebnis von Google bei einer Anfrage, die analog zu derjenigen im Kinderbuch gestellt wurde, ansieht (vgl. Abb. 4 und 5), dann scheint man das bereits erreicht zu haben!

Allerdings darf man sich von diesem netten Ergebnis nicht täuschen lassen: Die Suchmaschine hat natürlich die Frage nicht als solche analysiert oder gar „verstanden“, sondern die Eingabe lediglich als Buchstabenfolge aufgefasst und die Phrase „invented the phonograph“ in ihrem Index zu einem populären Dokument gefunden sowie dieses in der Rangliste ganz nach oben gesetzt bzw. ein Bild mit dem genau passenden umgebenden Text „Edison's first phonograph“ entdeckt. Insofern ist hier auch eine gehörige Portion Glück im Spiel gewesen; weniger primitive Fragen werden sich nicht so einfach ohne semantische Analyse beantworten lassen. Dennoch gibt dies eine aktuelle Forschungsrichtung vor: Von der Suchmaschine zur Antwortmaschine (vgl. Abschnitt 5).

Abb. 4: Google als Antwortmaschine



3 WIRTSCHAFTLICHE BEDEUTUNG VON SUCHTECHNOLOGIE

Die Online-Suche ist in kurzer Zeit zu einem eigenen Wirtschaftszweig, sogar zu einem „Big Business“ geworden. Im September 2007 war allein Google als börsennotiertes Unternehmen 174 Milliarden Dollar wert. Da es für Online-Unternehmen, inzwischen aber auch für viele klassische Firmen, entscheidend ist, dass ihre Webportale von potenziellen Kunden schnell gefunden werden, ist um die Suchmaschinen herum ein kompletter Wirtschaftszweig („Search Engine Optimization“) entstanden, der die Optimierung von Webseiten als Dienstleistung anbietet, damit diese eine bessere Rangposition in den Ergebnislisten erhalten. Die Grenze zwischen „Optimierung“ und „Manipulation“ ist allerdings fließend – Suchmaschinennutzer, und damit auch Suchmaschinenbetreiber,

haben natürlich kein Interesse daran, dass Websites in den Ergebnislisten auftauchen, die bezüglich der konkreten Suchanfrage wenig relevant sind. Deswegen, aber auch, um einen möglichen Vorteil vor der Konkurrenz nicht zu verspielen, halten die Suchmaschinenbetreiber ihre ständig verfeinerten Bewertungs- und Rankingkriterien geheim. Da Nutzer typischerweise nur den ersten paar Rangplätzen in den Suchergebnissen Beachtung schenken, ist der Kampf um die vorderen Plätze hart: Denn wenn einer aufsteigt, muss zwangsläufig ein Konkurrent absteigen.

Abb. 5: Google liefert auch das richtige Bild als Antwort



Da insbesondere reine Online-Unternehmen für einen steten Besucherstrom auf ihr Webportal sorgen müssen, kaufen sie oft von den Suchmaschinenbetreibern angebotene Schlüsselwörter. Gibt ein Suchmaschinennutzer einen solchen Begriff ein, so werden ihm, meist getrennt von den eigentlichen Suchresultaten, auch Links zu dem entsprechenden Unternehmen angezeigt. Eine andere Geschäftsmöglichkeit im Bereich der Online-Suche besteht darin, mit einer der großen Suchmaschinen wie Google oder Yahoo eine gewinnbringende Partnerschaft einzugehen: Besitzt man eine Website, die häufig besucht wird, dann kann man auf seinem eigenen Portal „Sponsored Links“ oder Werbeflächen vermieten, die durch die Suchmaschinenbetreiber mit zum Kontext der jeweiligen Webseite passenden Anzeigen gefüllt werden. Näheres zur Symbiose zwischen Online-Unternehmen und Suchmaschinen findet sich im Beitrag von Michael Buschmann in diesem Band.

Während sich die allgemeine Suche vor allem durch Werbeeinnahmen finanziert, ist dies bei dem sich in den letzten Jahren herausgebildeten „Enterprise Search“ nicht der Fall, weshalb hier auch nicht Google & Co., sondern andere Anbieter führend sind. Bei diesen Unternehmenssuchmaschinen werden zwar ähnliche technische Prinzipien

wie bei den klassischen Suchmaschinen angewendet, jedoch wird in besonderem Maße auf die Bedürfnisse von Unternehmen Rücksicht genommen. Es handelt sich also um eine Dienstleistung für Firmen, die diese entweder im Rahmen ihrer Wissensmanagement-Aktivitäten im eigenen Intranet oder beispielsweise als Marktforschungs-Tool zur Analyse von Wettbewerberseiten, Kunden-Foren und relevanten Blogs einsetzen. Auch die sogenannten „vertikalen“ Suchmaschinen stellen ein wirtschaftlich interessantes Marktsegment dar (siehe dazu den Beitrag von Stefan Wess und Jasmin Franz in diesem Band). Im Unterschied zu den universellen Suchmaschinen sind diese auf bestimmte Sachgebiete spezialisiert und daher in ihrem Umfang beschränkt. Sie können aufgrund ihrer Themenspezifität aber mit taxonomischen Hierarchien und spezifischen Wissensmodellen operieren, was ihre Suchergebnisse zwar weniger vollständig, aber dafür um ein Vielfaches präziser macht.

Die Bedeutung von eigenständigen Dienstleistungen, die auf Suchergebnissen aufbauen, dürfte ebenso wie das Marktpotenzial für das Enterprise Search und für die vertikalen Suchmaschinen in nächster Zeit deutlich zunehmen und stellt somit auch eine Chance für Unternehmen dar, die über spezifisches Know-how verfügen und sich damit nicht von vornherein der erdrückenden Konkurrenz der Marktführer der allgemeinen Online-Suche ausgesetzt sehen.

4 MACHT UND VERANTWORTUNG DER SUCHMASCHINEN

Der allgemeine Suchmaschinenmarkt ist von einem Oligopol, in Teilbereichen sogar von einer monopolartigen Struktur geprägt. Nach einem rasanten Wachstum in den letzten Jahren dominierte im September 2007 Google mit einem Marktanteil von 89,3 Prozent klar den deutschen Markt der Online-Suche; Yahoo kommt in Deutschland lediglich auf ca. 3 Prozent, T-Online auf rund 2 Prozent und MSN auf 1,6 Prozent.² Andere Suchanbieter, wie beispielsweise AOL, spielen praktisch keine Rolle mehr. In den USA ist das Bild nur wenig differenzierter: Google lag im September 2007 bei 57 Prozent, Yahoo bei ca. 24 Prozent und MSN bei ca. 10 Prozent.³ Es liegt also (praktisch weltweit, wenn man von China absieht, das hier aber auch aus anderen Gründen eine Sonderrolle spielt) eine marktbeherrschende Stellung einiger weniger US-Anbieter vor. Wie Marcel Machill et al. in diesem Band bemerken, sind zudem die Markteintrittschancen für neue Suchmaschinenwettbewerber gering, da der Betrieb der Infrastruktur einer leistungsfähigen universellen Suchmaschine mittlerweile hohe Kosten verursacht und die drei großen Marktführer nicht nur einen technologischen Vorsprung haben, sondern jährlich allein hunderte von Millionen Dollar für Forschung und Entwicklung ausgeben, um ihre Technologie weiterzuentwickeln und diesen Vorsprung auszubauen.

² WebHits 2007.

³ Search Engine Watch 2007.

Abb. 6: Stern-Titel 21/2006: „Weltmacht Google“



Suchmaschinen nehmen eine Schlüsselfunktion bei der Selektion und Vermittlung von Informationen im Internet ein. Gerade angesichts des Oligopols bzw. Quasi-Monopols kommt ihnen dadurch – ob sie es nun wollen oder nicht – eine bedeutende gesellschaftliche Verantwortung zu, der sie sich nicht einfach entziehen können. Tatsächlich wird die Auswahl von Inhalten im täglich wachsenden Informationsstrom in immer stärkerem Maße nicht mehr durch klassische „Gatekeeper“ wie Journalisten, sondern durch Suchmaschinen als digitale und algorithmische Pendanten vorgenommen, wie Machill et al. in ihrem Beitrag genauer ausführen und kritisch diskutieren. Ergänzend kommt hinzu, dass auch herkömmliche Medien ihre Information zunehmend aus dem Internet beziehen und Suchmaschinen natürlich auch für Journalisten den primären Zugang zum Internet darstellen.

Die Konzentration der Selektionsfunktion in wenigen Händen birgt grundsätzlich das Risiko des Missbrauchs, sodass sich – zumindest dann, wenn man Suchmaschinen nicht nur als reine Vermittler, sondern als Medien eigener Qualität, vergleichbar den Zeitungen, dem Rundfunk und dem Fernsehen, ansieht – die Frage nach der Notwendigkeit

einer staatlichen Regulierung stellt, die eventuell auch über die anderweitig bereits regulierten Aspekte des Jugend- und Urheberrechtsschutzes hinausgeht. Da die Rankingalgorithmen anfällig für externe Manipulation durch „Search Engine Optimizer“ sind, ist auch in Hinsicht auf das, was Suchmaschinen als relevant einstufen und vorne in den Ergebnislisten präsentieren, prinzipiell Vorsicht geboten. In diesem Kontext mahnt Rudi Schmiede in seinem Beitrag erhöhte Anstrengungen bei der Ausbildung von Medienkompetenz an, um die Urteilskraft des Nutzers zu stärken und auch, um auf dem Weg in die „Google-Gesellschaft“ erkennbaren Tendenzen zur Umwertung von Wissen entgegenzuwirken, wo an die Stelle von „wahr“ oder „falsch“ nun Kriterien der Zugänglichkeit wie „digitalisiert“ oder „nicht digitalisiert“ beziehungsweise gar „ergoogelbar“ oder „nicht ergoogelbar“ treten.

Die Monopolisierung eines gesellschaftlich und wirtschaftlich wichtigen Dienstes bei wenigen Anbietern wirft aber auch noch andere Fragen auf. Haben etwa einzelne Personen oder Firmen ein Recht darauf, in den Ergebnislisten von Suchmaschinen zu erscheinen? Oder können Suchmaschinen willkürlich – beispielsweise aus wirtschaftlichem Eigeninteresse – einige von ihnen boykottieren? Auch könnten Suchmaschinen mittels Data-Mining der Suchanfragen ein beträchtliches Wissen über möglicherweise wirtschaftlich relevante Aspekte (zum Beispiel Patentsuche, Aktienkurse, Trends) gewinnen und damit einen (unfairen?) Informationsvorsprung vor anderen Akteuren auf dem Markt erlangen.

Und wie sieht es mit den Nutzerprofilen (eingegebene Stichworte, Suchkontext, IP-Adresse etc.) aus, die die Suchmaschinenbetreiber erstellen? Aus Sicht der Betreiber kann damit sowohl die Relevanz der Suchergebnisse verbessert werden als auch zielgerichteter geworben werden. Allerdings unterliegen die Marktführer als außereuropäische Unternehmen nicht direkt den relativ strengen europäischen Datenschutznormen; ausländische Behörden könnten Zugriff auf die gesammelten Nutzungsdaten von Bürgern anderer Staaten erlangen. Zu den kostenlosen Suchdiensten merkt Rainer Kuhlen in seinem Beitrag daher an, dass das Bezahlen mit anderer Währung als mit Geld – gemeint ist das Bezahlen mit persönlichen Daten oder dem Offenlegen von Interessen und Verhaltensformen – vielleicht letztendlich sogar teurer kommen kann.

Einen Aspekt, den Kuhlen auch eher ambivalent sieht und in seinem Beitrag kritisch diskutiert, stellt das Bemühen Googles und großer Verlagsallianzen dar, Wissensbestände aus Büchern und ganzen Bibliotheken zu digitalisieren. Indem dies in Form eines Retaildienstes dann direkt den Kunden und Nutzern angeboten wird, erodieren die Leistungen der klassischen Mittler wie Bibliotheken, die bislang nicht gewinnorientiert arbeiten, aber als Teil ihrer öffentlichen Dienstleistung auch für eine Qualitätssicherung sorgen. Da „Information in elektronischen Umgebungen frei sein will“, Wirtschaftsunternehmen sich aber an Gewinnerwartungen orientieren, sei laut Kuhlen jedenfalls nicht zu erwarten, dass durch diese Art der Digitalisierung des Wissens ein Paradies der Wis-

sengesellschaft entstehen wird. Im Gegenteil: Die Informationsversorgung in Bildung und Wissenschaft könnte durch die Auslagerung in derartige Unternehmen zu einem Gut werden, das stückweise vom eigentlichen Produzenten wieder teuer zurückgekauft werden muss.

5 DIE SUCHMASCHINEN VON MORGEN

Bald werden nicht nur Texte, Fotos und Videos den Weg ins Internet finden, sondern auch ganz andere Daten wie etwa eingescannte Bücher, medizinische Datenbanken sowie viele von Sensoren in Realzeit generierte Daten zum Beispiel über die Umwelt, den Zustand von Produkten, den momentanen Ort von Gegenständen und Personen oder sogar über die eigene Gesundheit. Die Suche in solchen semantisch komplexen Datenbereichen erfordert „intelligente“ Mechanismen, die weit über heutige stichwortbasierte Verfahren hinausgehen, damit aus den Daten nützliche Information sowie letztlich Wissen und Wert entstehen.

Einige Forschungsrichtungen zeichnen sich unmittelbar ab. Da Suchmaschinen schon bald auch bequem auf mobilen Geräten wie Mobiltelefonen oder digitalen Assistenten („PDAs“) benutzt werden sollen, müssen die Suchmaschinen dann möglichst ohne langwierige oder explizite Texteingabe funktionieren. Indem der Ort und der Kontext des Nutzers erfasst werden, kann oft auf den vermutlich relevanten Diskursbereich geschlossen werden, sodass auch eine automatische Erkennung gesprochener Sprache und sogar ein inhaltliches Verständnis von (zumindest einfachen) Benutzerwünschen und -fragen erleichtert wird. Als Ausgabe ist in Situationen unterwegs typischerweise nicht ein langes Dokument gewünscht, sondern eine kurze Antwort, die für eine momentane Entscheidung genügt. Für solche Fälle bietet sich auch eine Sprachausgabe an. Wolfgang Wahlster geht in seinem Beitrag unter anderem auf diese Aspekte ein, die eine Suchmaschine zu einer „Antwortmaschine“ mutieren lassen.

Auch die Individualisierung und Personalisierung der Suche ist ein vielversprechendes Forschungsgebiet. Denn wenn eine Suchmaschine die Suchhistorie, den Dialogkontext, die Interessensgebiete, den Bildungsstand, die Fremdsprachenkenntnisse, die Vorlieben oder beispielsweise auch die Krankheitsgeschichte des Nutzers kennt (beziehungsweise im Laufe der Zeit kennen lernt), dann kann sie ihn natürlich viel gezielter mit relevanter Information versorgen. Ein prinzipielles Problem stellt dabei allerdings der Schutz der hierfür notwendigen oder „mitgehört“ privaten und intimen Informationen vor dem Zugriff Unbefugter dar.

Einen weiteren, schon lange gehegten, aber wohl nicht einfach oder schnell zu erfüllenden Wunsch stellt die Suche in ganzen Bedeutungsklassen anstelle der relativ dazu eher primitiven Suche identischer Dinge („letter-by-letter, pixel-by-pixel“) dar, sodass etwa ähnliche Bilder zu einem vorhandenen Referenzbild gefunden oder Analogien in der Struktur von Proteinen automatisch entdeckt werden können.

Schließlich seien noch das „Semantic Web“ und das „Web 3.0“ genannt, auf denen derzeit große Erwartungen ruhen. Da bisher im Web Dokumente nicht maschinell verstehbar sind, ist es schwierig, diese inhaltlich zu kategorisieren und entsprechenden Suchanfragen gut zuzuordnen – dies geschieht bisher nur oberflächlich auf der Basis von Stichworten. Das semantische Web basiert auf der inhaltlichen Beschreibung digitaler Dokumente in Form von Ontologien, die eine standardisierte Begrifflichkeit zur Beschreibung digitaler Inhalte bereitstellen und damit eine maschinell verstehbare Semantik realisieren. Damit sollte ein viel „tiefer gehendes“ Verständnis der Textinhalte möglich sein und somit auch eine präzisere Antwort auf eine Suchanfrage gegeben werden können. Die semantische Annotation von Dokumenten ist allerdings mühsam. Die diesbezügliche Hoffnung beim Web 3.0 beruht auf seiner saloppen Definition „Web 3.0 = Semantisches Web + Web 2.0“: Man möchte also das Prinzip der Benutzerpartizipation des Web 2.0 als Massenbewegung für den Aufbau des semantischen Webs nutzen. In welchem Maße dies gelingen wird, kann erst die Zukunft zeigen. Der Beitrag von Wahlster geht auch hierauf näher ein.

6 LEBEN IM ZEITALTER DES „INSTANT SEARCH“

Suchmaschinen, im Zusammenspiel mit Systemen wie Wikipedia, Online-Auskunftssystemen, Social-Networking-Plattformen sowie mobilem Informationszugang mittels Handy oder digitalen Assistenten, ändern unsere Lebensgewohnheiten schnell und radikal. Schon 2005 drückte dies Joanna Glasner in „Wired“ auf nette Art so aus: „Google’s just seven years old, but it’s hard to imagine life before instant search“.⁴

Tatsächlich müssen sich Wissenschaftler ja kaum noch in eine Bibliothek bemühen; nicht nur Bibliothekskataloge, sondern auch Fachzeitschriften sind online, und Google selbst dient oft als Ersatz für beides – „the information of the world at your finger-tips“. Ähnliches gilt für Lexika und Wörterbücher: Wikipedia (vielleicht noch ergänzt durch den Online-Duden) ist schneller, aktueller und oft auch umfassender – und Google leitet einen gerne mit einem einzigen Klick dort hin. Fremdsprachenwörterbücher haben ebenfalls ausgedient; die durch Tausende freiwilliger Nutzer gepflegten Online-Wörterbücher erfüllen den Zweck genauso gut. Dem guten alten Taschenrechner haben wir neulich, als der Funkmaus der Strom ausging, die Batterien geraubt und dann nie wieder ersetzt, denn eine kleine Rechenaufgabe ist bei Google ja mindestens genauso schnell eingetippt und sofort gelöst, und das Ergebnis kann zudem mit Copy & Paste sofort weiterverarbeitet werden. Überhaupt Copy & Paste: Schnell zusammengegoogelte Dokumente zu einem Stichwort liefern das Rohmaterial für den „eigenen“ Text, bei dem oft nur noch die Nahtstellen etwas zurechtgeschliffen werden. Die eigenen Veröffentlichungen sind vom Google Cache aus schneller zugreifbar als vom Dateisystem des PCs, und die E-Mail wird bei Google Mail sicherer gespeichert als in der lokalen Mailbox. Filme zum Fotografieren, Terminkalender aus Papier oder Straßenkarten hat man seit Jahren nicht

⁴ Glasner 2005.

mehr gekauft. Für die beabsichtigte Fahrt von „hier“ nach „dort“ liefert Google – als hätte die Suchmaschine meinen Wunsch geahnt! – bei Eingabe der beiden Orte direkt die passenden Zugverbindungen; fahre ich mit dem Auto, erhalte ich die Fahrstrecke auf einer Straßenkarte eingezeichnet. Interessiert mich der Publikationsindex eines mehr oder auch eines weniger geschätzten Kollegen, dann lasse ich mir seine Zitationszahlen von Google Scholar oder dem darauf aufbauenden Online-Tool mit dem sinnigen Namen „Publish or Perish“ in null Komma nichts berechnen. Von Bewerbern erfährt man zumindest den Sportklub und den Titel der Seminararbeit, Bewerberinnen haben meist auch gleich noch die ganzen Urlaubsfotos der letzten Jahre online. Und will man trotz der vielen nie abonnierten Newsletter wissen, was eine „echte“ Zeitung schreibt, kommt man mit Google News direkt zu den relevanten Online-Artikeln.

Eine schöne neue Welt? In mancher Hinsicht sicherlich, andererseits gehen mit ihr auch einige alte Tugenden verloren, die Zeit und Muße benötigen oder die nur aufgrund früherer Unzulänglichkeiten der analogen, dinglichen Welt entstanden sind. Kopfrechnen, Rechtschreibung, Geduld und möglicherweise auch Bescheidenheit gehören vielleicht dazu. Aber noch ist es für eine Bilanz zu früh, denn die Entwicklung hin zur totalen Digitalisierung unseres Wissens (oder gar unserer Welterfahrung) ist noch lange nicht abgeschlossen. Und so wie vor zwanzig Jahren kaum jemand Webbrowser, Suchmaschinen, digitale Musik und Geschäftsmodelle für Klingeltöne vermisst oder auch nur vorausgeahnt hatte, so mag die Zukunft noch für viele digitale Überraschungen und diverse darauf aufbauende Nutzungsarten mit ungeahnten Geschäftsmöglichkeiten gut sein. Vielleicht aber auch für so manche Herausforderung und allerhand Streit – denn wer darf die Welt digitalisieren und wer tut es schließlich? Und wem gehören die Daten und wer profitiert letztendlich von dem daraus destillierten Wissen?

Wie auch immer die Zukunft aussehen wird: Ohne Suchmaschinen oder ähnliche Instrumente werden wir wohl kaum mehr auskommen können, denn „mit der Speicherung von Wissen in digitalen Dokumenten sind wir zum ersten Mal seit Erfindung der Schrift darauf angewiesen, Maschinen zu besitzen, die in der Lage sind, uns Zugang zum gespeicherten Wissen zu verschaffen.“⁵

7 LITERATUR

Childcraft 1964

Childcraft International: *Childcraft Encyclopedia, Volume 6: How Things Change*. Field Enterprises Educational Corporation, 1964.

Glasner 2005

Glasner, Joanna: „Futurists Pick Top Tech Trends.“ In: *Wired*, 25.10.2005. URL: <http://www.wired.com/techbiz/startups/news/2005/10/69138?currentPage=all> [Stand: 12.12.2007].

⁵ Plaß 2005.

Plaß 2005

Plaß, Christine: „Das große Vergessen. Datenschwund im digitalen Zeitalter.“ In: Lehman, Kai/Schetsche, Michael (Hrsg.): *Die Google-Gesellschaft. Vom digitalen Wandel des Wissens*. Bielefeld: transcript, 2005, S. 41-46.

Search Engine Watch 2007

Burns, Enid: „U.S. Search Engine Rankings, September 2007.“ In: *Search Engine Watch*, 20.11.2007. URL: <http://searchenginewatch.com/showPage.html?page=3627654> [Stand: 12.12.2007].

WebHits 2007

WebHits: *Web-Barometer*. URL: www.webhits.de [Stand: September 2007].

> SUCHMASCHINENFORSCHUNG: ÜBERBLICK UND SYSTEMATISIERUNG EINES INTERDISZIPLINÄREN FORSCHUNGSFELDES¹

MARCEL MACHILL/MARKUS BEILER/MARTIN ZENKER

1 EINLEITUNG

Suchmaschinen sind aus dem Internet nicht mehr wegzudenken. Das Wachstum der Netzinhalte, begünstigt durch die niedrigen kommunikativen Zugangsbarrieren, machen die Selektion relevanter Inhalte, das Trennen des Wichtigen vom Unwichtigen, zur Notwendigkeit. Suchmaschinen übernehmen diese Selektions- und Vermittlungsfunktion an der Schnittstelle zwischen öffentlicher und individueller Kommunikation. Sie ermöglichen in vielen Fällen erst den Zugang zu ansonsten schwer oder gar nicht zugänglichen Informationen und erweitern somit den Horizont der rein menschlichen Suche.² Suchmaschinen liefern nicht automatisch die ‚richtigen‘ Antworten auf Suchanfragen, sie reduzieren die Komplexität des Webs auf eine überschaubare Zahl von Seiten, von denen aus eine Nutzerin oder ein Nutzer in die Lage versetzt wird, ein für sie bzw. ihn relevantes Ergebnis zu finden. Sie üben somit eine den klassischen Gatekeepern vergleichbare Funktion aus.³

Entsprechend ihrer Schlüsselrolle sind Suchmaschinen nach E-Mails die am zweithäufigsten genutzte Internet-Anwendung: 75 Prozent aller deutschen Onliner nutzen mindestens einmal pro Woche eine Suchmaschine (78 Prozent E-Mails).⁴ In den USA greifen 84 Prozent der Internetnutzer regelmäßig auf Suchmaschinen zurück, 56 Prozent sogar täglich.⁵

Prägend für den Suchmaschinenmarkt ist das Oligopol der drei US-Suchmaschinenbetreiber Google, Yahoo und MSN, die weltweit den Großteil des Suchmaschinenmarktes unter sich aufteilen. Google nimmt in diesem Triumvirat die Spitzenposition ein, wofür der Begriff „Googlepol“ steht.⁶ Mit einem Markt- bzw. Suchanteil von 57 Prozent dominiert Google den deutschen Markt der Online-Suche. MSN kommt auf elf Prozent, Yahoo auf zehn und die Suchfunktion von AOL auf neun Prozent. Die übrigen Anbieter zusammen erreichen nur einen Marktanteil von 13 Prozent (Dezember 2005).⁷ In den USA konnte Google innerhalb eines Jahres seinen Suchanteil um sechs Prozent auf knapp

¹ Bei dem nachfolgenden Beitrag handelt es sich um einen gekürzten Zweitabdruck des gleichnamigen Kapitels aus dem Buch *Die Macht der Suchmaschinen / The Power of Search Engines*, herausgegeben von Marcel Machill und Markus Beiler (Köln: Halem 2007). Wir danken dem Herbert von Halem Verlag für die freundliche Genehmigung.

² Rieder 2005, S. 29f.

³ Machill/Neuberger/Schindler 2002, S. 15ff.; Wolling 2002.

⁴ Van Eimeren/Frees 2006, S. 406.

⁵ Fallows 2005.

⁶ Winterbauer 2003.

⁷ iBusiness 2006.

die Hälfte (49 Prozent) im Dezember 2005 vergrößern, während Yahoo bei 21 Prozent verharrte sowie MSN um drei auf elf Prozent und die sonstigen Anbieter um zwei auf 19 Prozent verloren.⁸ Daraus resultiert eine erhebliche Marktmacht dreier Anbieter.

Suchmaschinen sind keine „Maschinen“ im traditionellen Sinne. Nach Rieder können Suchmaschinen beschrieben werden als Software, die einen Index definierter Daten erstellt, der über Retrieval-Methoden erschließbar ist und einen bestimmten Anzeigemodus verwendet, um die Suchergebnisse zu präsentieren.⁹ Die Erstellung des Indexes erfolgt automatisch mittels Crawler (auch Robots bzw. Spider genannt), die kontinuierlich Internetseiten aufsuchen und die Inhalte der Seite in komprimierter Form speichern. Aus diesem Index generieren Suchmaschinen bei von Nutzern gestellten Suchanfragen die Ranglisten, indem sie alle indextierten Seiten auf ihre angenommene Relevanz in Bezug auf die Suchanfrage überprüfen und die zugehörigen Links entsprechend anordnen. Die Relevanz-Kriterien, die hierbei zum Einsatz kommen, werden als Firmengeheimnisse der Suchmaschinenunternehmen gehütet.

Das Phänomen der Suchmaschinen, wie wir sie heute kennen, ist jung. Der Vorläufer war im Jahr 1990 das Programm „Archie“, das in der Zeit vor dem World Wide Web unterschiedliche FTP-Server ansteuerte und Listen mit den dort verfügbaren Dokumenten erstellte, die von anderen Nutzern durchsucht werden konnten.¹⁰ Das erste Programm, das den Funktionen heutiger Suchmaschinen ähnlich war, war der „World Wide Web Wanderer“. Es nutzte bereits einen Robot, um Webseiten zu indexieren. Zeitgleich mit der Veröffentlichung der ersten massenkompatiblen Browser von Netscape und Microsoft Ende 1994 bzw. Anfang 1995 und dem damit einhergehenden sprunghaften Anstieg der Nutzung des World Wide Web nahm auch der Bedarf nach Orientierung im wachsenden Netz zu. Es entstanden die ersten großen Suchmaschinen und durchsuchbaren Kataloge, unter anderem Yahoo, Excite, Lycos, AltaVista und HotBot. Bereits in dieser Prä-Google-Ära wurde die Macht der Suchmaschinen ebenso wie die „Blindheit“ der Nutzer im Umgang mit ihnen thematisiert.¹¹ Aber erst mit dem Aufstieg Googles und der Vormachtstellung eines Internet-Konzerns erreichte die Diskussion um die Macht von Suchmaschinen breite Bevölkerungskreise.

Im Gegensatz zu der großen gesellschaftlichen Bedeutung und der Macht von Suchmaschinen weist die Forschung Lücken auf, die es aufzuarbeiten gilt. Dieser Beitrag will hierzu inhaltlich wie systematisch beitragen. Nachfolgend werden fünf Themenfelder herausgearbeitet, die wir für die künftige Suchmaschinenforschung aus interdisziplinärer Perspektive als zentral erachten und die einer künftigen Systematisierung der relevanten Forschung dienen kann.

⁸ Nielsen Netratings 2006.

⁹ Rieder 2005, S. 27f.

¹⁰ Vgl. im Folgenden Sonnenreich 1997.

¹¹ Winkler 1997.

2 SUCHMASCHINENPOLITIK UND -REGULIERUNG

Der Bedarf der medienpolitischen Debatte über Suchmaschinen ergibt sich zum einen aus der hohen Konzentration im Suchmaschinenmarkt und der damit verbundenen *Marktmacht* einiger weniger Anbieter und zum anderen aus inhaltlichen Problemen.

Die Marktmacht und damit die Notwendigkeit der Regulierung resultieren vor allem aus dem Oligopol auf dem Suchmaschinenmarkt. Die Konzentration der Selektionsfunktion in wenigen Händen birgt das potenzielle Risiko des Missbrauchs. Hinzu kommt, dass es auch unter wirtschaftlichen Aspekten ein Oligopol von Google, Yahoo und MSN gibt, der Werbemarkt für Suchmaschinen von diesen drei Anbietern fast vollständig dominiert wird und damit Abhängigkeiten von Werbenden gegenüber den Suchmaschinenbetreibern bestehen, insbesondere dann, wenn das Geschäftsmodell des Werbetreibenden auf dem Internet aufbaut (vgl. Abschnitt 3: Suchmaschinenökonomie).

Die Konzentration auf dem Suchmaschinenmarkt ist tatsächlich noch gravierender, da es in den letzten Jahren zu zahlreichen Firmenübernahmen kam und die Suchmaschinen darüber hinaus über Lieferverträge miteinander verknüpft sind.¹² So liefert Google normale Suchergebnisse an AOL, T-Online und Netscape sowie gesponserte Links an AOL, Lycos, Hotbot, Netscape und Ask. Bis Februar 2004 wurde aufgrund alter Verträge auch Konkurrent Yahoo versorgt, obwohl dieser bereits im Dezember 2002 den Indexdienstleister Inktomi erworben hatte. Von Yahoo erhalten normale Einträge die zur Unternehmensgruppe gehörenden Alltheweb und Altavista. Yahoo Search Marketing, das aus der im Juli 2003 erworbenen Vermarktungsfirma Overture hervorging, liefert bezahlte Einträge neben diesen beiden Suchmaschinen an Web.de. MSN Search, zuvor auch von Yahoo beliefert, stützt sich seit Januar 2005 auf einen eigenen Index und vermarktet in den USA seit Mai 2006 über Microsoft adCenter auch die Suchwortwerbung selber.¹³

Der Aspekt der Marktmacht wird noch offensichtlicher, wenn man das Verhalten der Nutzer hinzuzieht. So zeigen Studien aus Deutschland und den USA, dass nur rund die Hälfte aller Internetnutzer mehr als eine Suchmaschine verwendet.¹⁴ Dies resultiert in einer hohen gesellschaftlichen Verantwortung der Suchmaschinenbetreiber, die nicht allein durch den stark konzentrierten Markt zu gewährleisten ist. Denn der Anspruch der Suchmaschinenbetreiber, dem Nutzer relevante Ergebnisse zu liefern, kollidiert mit ihrem Geschäftsmodell, das eine Refinanzierung durch den Verkauf von Werbeflächen vorsieht.¹⁵

Die Marktmacht der drei führenden Suchmaschinenbetreiber ist auch vor dem Hintergrund zu diskutieren, ob ein Recht auf Auflistung in den Ergebnislisten besteht. Schulz, Held und Laudien haben festgestellt, dass die Anbieter von Inhalten in Deutschland keine rechtliche Handhabe besitzen, um eine Aufnahme in den Index einer Suchmaschine zu erlangen.¹⁶ Lediglich im Bereich der Suchwort-Vermarktung sei ein Anspruch eines Anbieters gegenüber einer Suchmaschine, sofern diese eine marktbeherrschende Stellung besitzt, denkbar.

¹² Bruce Clay 2006.

¹³ Beiler/Machill 2007.

¹⁴ Fallows 2005, S. 13f.; Machill/Neuberger/Schweiger/Wirth 2003, S. 149.

¹⁵ Van Couvering 2004; Vine 2004.

¹⁶ Schulz/Held/Laudien 2005, S. 79ff.

Die Marktmacht ist medienrechtlich in Deutschland bisher nicht reguliert worden. Ein Grund ist unter anderem die nicht einfache Einordnung von Suchmaschinen in den bestehenden Rechtsrahmen.¹⁷ Gleichwohl muss darauf hingewiesen werden, dass eine Konzentration des beschriebenen Ausmaßes insgesamt und eine marktbeherrschende Stellung wie die von Google bei anderen elektronischen Medien sowohl in Deutschland als auch auf dem Heimatmarkt USA nicht zulässig wäre. So vermutet der deutsche Rundfunkstaatsvertrag eine marktbeherrschende Stellung bei einem Zuschaueranteil von 30 bzw. 35 Prozent (bei weiteren medienrelevanten Aktivitäten) und sieht Gegenmaßnahmen vor. In den USA schreibt der Telecommunications Act bei Fernsehsendern eine Reichweitenbegrenzung von 35 Prozent vor. Entsprechende Regelungen für den Suchmaschinenmarkt gibt es nicht. Bislang greift in Deutschland nur das allgemeine Kartellrecht, das auf Unternehmenszusammenschlüsse begrenzt ist und inneres Wachstum nicht kontrollieren kann.¹⁸

Während die Konzentration auf dem Suchmaschinenmarkt noch unreguliert ist, gibt es bei *inhaltlichen Aspekten* bereits einige Ansätze. Gleichwohl wird die inhaltliche Regulierung dadurch erschwert, dass Suchmaschinen als Teil des Internets global agieren und in Ländern, in denen sie keine Infrastruktur betreiben, schwer zu sanktionieren sind. Klassischer nationalstaatlicher Regulierung sind damit Grenzen gesetzt. In den USA beruht die inhaltliche Regulierung auf dem sich ausbildenden Fallrecht nach konkreten Rechtsstreitigkeiten. Dieses korrespondiert mit technischen Entwicklungen der Suchmaschinen und der Entstehung bestimmter Geschäftsmodelle. Die Gesetzgebung und Regulierungsbehörden in den USA beschäftigten sich parallel zu den Entscheidungen der Gerichte mit der Inhaltsregulierung von Suchmaschinen (zum Beispiel beim Jugendschutz), mit Fragen der Haftbarkeit von Suchmaschinen bei Urheberrechtsverstößen und dem Verbraucherschutz.¹⁹

In Deutschland ging die sich entwickelnde inhaltliche Regulierung von Suchmaschinen in erster Linie von der Fragestellung aus, wie der Jugendschutz gesichert werden kann. So besteht trotz steigender selbstregulatorischer Anstrengungen der Suchmaschinen die Gefahr, dass bei scheinbar harmlosen Suchbegriffen Internetseiten mit pornografischem oder gewalthaltigem Inhalt angezeigt werden. Dies ist unter anderem dem Umstand geschuldet, dass die Algorithmen der Suchmaschinen nicht die inhaltliche Bedeutung der aufgefundenen Seiten ‚verstehen‘.²⁰ Besondere Schwierigkeit bereiten den Suchmaschinen auch Bilder und ähnliche Elemente, die ca. 70 Prozent aller Inhalte im Netz ausmachen.²¹ So ist mittels gezielter Manipulation ein Platzieren einer Seite mit jugendgefährdendem Inhalt in den Ergebnislisten der Suchmaschinen möglich.²² Ob ein Foto beim Suchbegriff „Pflaume“ die Frucht eines Obstbaumes oder aber weibliche Genitalien abbildet, ist bei einer maschinellen Verarbeitung nicht zweifelsfrei zu ermitteln. Hier sind Suchmaschinen gefordert, verstärkt neben rein automatischer Inde-

¹⁷ Vgl. Schulz/Held/Laudien 2005.

¹⁸ Schulz/Held/Laudien 2005, S. 96.

¹⁹ Vgl. Gasser/Thurman 2007.

²⁰ Vgl. Machill/Neuberger/Schweiger/Wirth 2003, S. 27f.

²¹ Green 2000, S. 131.

²² Vgl. Machill/Neuberger/Schindler 2002.

xierung auch eine manuelle Qualitätskontrolle aufzubauen, die zumindest stichprobenartig prüfen und gezielt eingreifen kann.

In Deutschland hat sich ein Modell der regulierten Selbstregulierung herausgebildet.²³ So ermöglicht der 2003 in Kraft getretene Jugendmedienschutz-Staatsvertrag (JMStV) Einrichtungen Freiwilliger Selbstkontrolle auch für den Bereich des Internets. Im April 2005 haben die wichtigsten deutschen Suchmaschinenbetreiber unter dem Dach der Freiwilligen Selbstkontrolle Multimedia-Diensteanbieter (FSM) eine eigene Selbstkontrolle vereinbart, nachdem sie sich im Dezember 2004 bereits auf einen Verhaltenskodex verständigt hatten. Damit wurde eine entsprechende Initiative der Bertelsmann Stiftung aufgegriffen.²⁴ Der Code of Conduct regelt den Umgang mit jugendgefährdenden oder eindeutig illegalen Internetseiten, die zum Beispiel Volksverhetzung, die Ausschwitzlüge oder Kinderpornografie verbreiten.²⁵ Maßnahmen sind der Ausschluss entsprechender Seiten oder der Einsatz von Familienfiltern. Bei Beanstandungen ist die Beschwerdestelle der FSM anzurufen. Je nach Schwere des Verstoßes sind Sanktionen vorgesehen.²⁶ Daneben haben sich die Suchmaschinenbetreiber dazu verpflichtet, werbliche Suchergebnisse angemessen zu kennzeichnen (siehe dazu auch Abschnitt 3: Suchmaschinenökonomie) und bei der Erfassung und Verwendung von Nutzerdaten Zurückhaltung zu üben.

Das Ausblenden von Inhalten bei Suchmaschinen muss in Ländern mit einem restriktiven politischen System kritisch betrachtet werden, wenn systemkritische Inhalte hierdurch zensiert werden.²⁷ In diesem Zusammenhang werden vor allem Google und Yahoo für ihr Engagement in China angegriffen, da sie Vorgaben der Pekinger Zentralregierung teilweise gewissenhafter umsetzen als etwa der chinesische Suchmaschinenmarktführer Baidu.ch.²⁸ Google wehrt sich gegen die Vorwürfe und verweist darauf, dass eine unzensurierte Google-Seite mit chinesischen Angeboten außerhalb Chinas weiterhin gehostet wird.²⁹ Um mögliche Gesetzesentscheidungen seitens der USA, die Handlungen von US-Unternehmen, die die Entwicklung von Demokratie behindern, im Ausland zu untersagen, ist der Yahoo-Service in China mittlerweile in eine chinesische Tochterfirma ausgegliedert.³⁰

Suchmaschinen zeichnen das Verhalten ihrer Nutzer auf, speichern IP-Adressen und Suchanfragen, um Nutzerprofile zu erstellen. Damit soll sowohl die Relevanz der Suchergebnisse verbessert werden als auch zielgerichteter geworben werden.³¹ Diese Sammlung von Nutzerdaten ist aus mehreren Gründen problematisch. Zum einen haben die Nutzer der Suchmaschine keine Kenntnis darüber, in welchem Umfang ihr Suchverhalten aufgezeichnet und weiterverwertet wird. Zum anderen gibt es für Nutzer keinerlei Möglichkeit, einen Einblick in die über sie gesammelten Daten zu erhalten, wie es ihnen eigentlich aus Gründen der informationellen Selbstbestimmung zustehen sollte.

²³ Vgl. Beiler/Machill 2007.

²⁴ Vgl. FSM 2005; Welp 2003.

²⁵ FSM 2004a.

²⁶ FSM 2004b.

²⁷ Vgl. Schulz/Held 2007.

²⁸ Spiegel Online 2006a; Heise online 2006a; Rötzer 2006; CNNIC 2005.

²⁹ Vgl. Beiler/Zenker 2006, S. 13f.

³⁰ Rötzer 2006.

³¹ Vgl. Sullivan 2006b.

Problematisch wird der Aspekt der Nutzerdatensammlung auch dadurch, dass die Daten europäischer Nutzer von Google, Yahoo, MSN oder anderen außereuropäischen Suchmaschinen nicht auf den Servern in der Europäischen Union, sondern in den USA gespeichert werden. Wenn nun die Suchmaschinen in den USA Nutzerdaten an die dortigen Behörden herausgeben, beispielsweise aufgrund des „Patriot Act“, können somit ausländische Behörden Zugriff auf die gesammelten Nutzungsdaten von Bürgern anderer Staaten erlangen.³² Dieses Problem muss von der Medienpolitik auf transnationaler Ebene geklärt werden.³³ Vor der Perspektive des Web 2.0, dem Versuch der Personalisierung und Netzwerkbildung im Internet, wird diese Problematik an Brisanz gewinnen.

Auch mit Copyrightaspekten muss sich die Medienregulierung beschäftigen. Suchmaschinen indizieren quasi automatisch auch Inhalte, die dem Copyright unterliegen, sofern dies nicht technisch unterbunden ist.³⁴ Da die Suchmaschinenbetreiber durch das Platzen von Werbung sehr gut am Verweis auf fremdes Eigentum verdienen, haben die Urheber des geistigen Eigentums teilweise ein Interesse daran, an diesen Werbeeinnahmen – die ohne ihre Inhalte nicht im gleichen Maße fließen – zu partizipieren. Besonders deutlich wird dies bei Nachrichtensuchmaschinen wie Google News, bei denen der Suchmaschinenbetreiber auf Inhalte Dritter verweist, ohne die die Nachrichtensuche wertlos wäre.³⁵

Auch Projekte wie Google Books, die die Digitalisierung der Wissensbestände aus Büchern zum Ziel haben, müssen vor diesem Hintergrund betrachtet werden. Einerseits wird die Digitalisierung von Büchern, deren Copyright abgelaufen ist, begrüßt. Andererseits fürchten Verlage, dass über die Suchmaschine ihre Inhalte kostenlos im Netz erhältlich sein könnten und die Rechteinhaber leer ausgehen. Dementsprechend werden seitens der Verlage bereits Konkurrenzprodukte zu den Suchmaschinen entwickelt.³⁶

Insgesamt ist festzuhalten, dass sich in Deutschland ansatzweise ein Governancesystem entwickelt hat, das Teilantworten auf die Vielzahl der genannten Regelungsaspekte von Suchmaschinen geliefert hat, die insbesondere im Jugendschutz und in bezahlten Links zu finden sind. Andere inhaltliche Aspekte und infrastrukturelle Probleme der Konzentration sind dagegen noch nicht angegangen. Daher muss das System der Kontrolle von Markt- und Meinungsmacht auf den Bereich Suchmaschinen erweitert werden. Maßnahmen könnten unabhängige, von gesellschaftlich relevanten Gruppen besetzte Beiräte sein, die zum Beispiel darauf achten, dass keine Inhalteanbieter im Zugang zur Suchmaschine diskriminiert werden. Eine flankierende Maßnahme wäre darüber hinaus die Pflicht für Suchmaschinen, ihr unternehmerisches und publizistisches Handeln öffentlich in regelmäßigen Berichten zu rechtfertigen.³⁷

Neben solchen Regelungsoptionen, die infolge der Globalität des Internets nur schwer durchsetzbar sind, zeichnen sich derzeit drei Pendanten zum Oligopol der drei privaten US-Suchmaschinen ab. Erstens wird über eine zentrale öffentlich-rechtliche

³² Heise online 2006f.

³³ Vgl. Rotenberg 2007.

³⁴ Vgl. Crowell 2005.

³⁵ Vgl. Keith 2007.

³⁶ Vgl. Köver 2006.

³⁷ Vgl. dazu allgemein Trappel/Meier/Schrape/Wölk 2002.

Suchmaschine diskutiert,³⁸ die Suchergebnisse ohne kommerziellen Einfluss darstellen würde,³⁹ jedoch einen hohen Finanzierungsbedarf hätte⁴⁰. Zweitens wäre eine Förderung freier Suchmaschinen denkbar, die sich auf eine dezentrale Open-Source-Infrastruktur stützen.⁴¹ Diesem Ziel hat sich der Verein zur Förderung der Suchmaschinen-Technologie und des freien Wissenszugangs (SuMa-eV) verschrieben. Drittens hat der deutsch-französische Ministerrat im April 2005 die Förderung der Entwicklung einer europäischen Suchmaschine namens Quaero beschlossen (siehe dazu Abschnitt 5: Technik und Qualität von Suchmaschinen). Damit sollte auch eine nicht wieder aufholbare Vorherrschaft der USA bei der Überlieferung kulturellen Wissens verhindert werden.⁴² Aufgrund von Differenzen in der Grundausrichtung wird das Projekt ab 2007 voraussichtlich in zwei Einzelprojekten – der Entwicklung einer Konkurrenzsuchmaschine zu Google sowie Grundlagenforschung zur semantischen Suche – fortgeführt.⁴³

3 SUCHMASCHINENÖKONOMIE

Suchmaschinen sind ein „Big Business“: Legt man den Börsenwert zugrunde, so ist Google das „wertvollste Medienunternehmen der Welt“.⁴⁴ Im Oktober 2006 übersprang der Börsenwert von Google die Marke von 150 Milliarden US-Dollar. Damit war der Suchmaschinenbetreiber Google bereits wertvoller als das Technologie-Schergewicht IBM, das es auf eine Marktkapitalisierung von 140 Milliarden US-Dollar brachte.⁴⁵ Der Sprung beim Aktienkurs war vor allem den außergewöhnlich guten Ergebnissen Googles im Geschäftsjahr 2006 zu verdanken. Allein im dritten Quartal stieg der Konzerngewinn um 92 Prozent von 381 Millionen US-Dollar auf 733 Millionen US-Dollar an.

2005 lag der Konzernumsatz Googles bei 6,1 Milliarden US-Dollar,⁴⁶ Yahoo erreichte einen Umsatz von 5,3 Milliarden US-Dollar.⁴⁷ Während Google einen Gewinn von 1,47 Milliarden US-Dollar erzielte – über fünf Dollar pro Aktie –, erwirtschaftete Yahoo im gleichen Geschäftsjahr einen Gewinn von 1,9 Milliarden US-Dollar. Damit war 2005 für beide Unternehmen das bisher ertragreichste Jahr ihrer Geschichte. Zum Vergleich: Der TimeWarner-Konzern, größtes Medienunternehmen der Welt, brachte es im gleichen Geschäftsjahr bei einem sehr viel höheren Umsatz von 43,65 Milliarden US-Dollar ‚nur‘ auf einen Konzerngewinn von 2,92 Milliarden US-Dollar.⁴⁸

Zu beachten ist, dass Suchmaschinen sehr stark vom Werbemarkt abhängig sind. Während der Anteil der Werbeeinnahmen 2005 bei Google bei 98,8 Prozent des Gesamtumsatzes lag,⁴⁹ machten die Werbeeinnahmen am Gesamtumsatz im stark diversifizierten Medienkonzern TimeWarner lediglich 17,4 Prozent aus, wenngleich dieser in

³⁸ Vgl. Sadrozinski 2007.

³⁹ Plog 2006.

⁴⁰ Heise online 2004.

⁴¹ Ebenda.

⁴² Heise online 2005.

⁴³ Stöcker/Darnbeck 2006.

⁴⁴ Schotzger 2005.

⁴⁵ Spiegel Online 2006b.

⁴⁶ Google 2006a.

⁴⁷ Yahoo 2006.

⁴⁸ TimeWarner 2006.

⁴⁹ Google 2006a, S. 48.

absoluten Zahlen 1,5 Milliarden US-Dollar über dem Googles lag.⁵⁰ Diese einseitige Ausrichtung des Geschäfts auf den Werbeverkauf hat für Suchmaschinen zur Folge, stetig neue Nutzer gewinnen und hohe Zugriffszahlen generieren zu müssen, um dadurch die Attraktivität als Werbeplatz zu erhöhen.

Auch vor diesem Hintergrund ist beispielsweise die Übernahme des Video-Portals Youtube.com durch Google für 1,65 Milliarden US-Dollar zu sehen. Während YouTube.com lediglich 25 Millionen US-Dollar Werbeumsatz besitzt, sind vor allem die 30 bis 40 Millionen Nutzer weltweit für Google interessant. Allein in Deutschland nutzten im Mai 2006 3,2 Millionen Menschen das kostenlose Videoportal.⁵¹ Die ebenfalls stark frequentierte Community-Website MySpace.com wurde von Google durch die Zahlung von 900 Millionen US-Dollar für mehrere Jahre als Werbepartner rekrutiert.⁵² Ein weiterer Schritt zur Erschließung von neuen Absatzmärkten für Werbung ist Googles AdSense-Programm, bei dem private Nutzer von Google gelieferte, zum Inhalt der Seite passende Anzeigen auf ihren privaten Homepages anzeigen lassen können und an den Einnahmen aus dieser Werbung beteiligt werden.⁵³

Einen vollkommen neuen Schritt in der Erschließung neuer Werbepotenziale stellt eine Kooperation Googles mit mehreren US-Tageszeitungen dar. Die Zeitungsverlage wollen die Werbeplätze in Print- und Online-Ausgaben durch Google vermarkten lassen, um auf diese Weise neue Werbepartner zu generieren. Google gewinnt durch diesen Deal neue Absatzmöglichkeiten.⁵⁴

Nicht immer funktioniert die Zusammenarbeit zwischen Suchmaschinenbetreiber und Inhalteanbieter derart reibungslos. Da eine Vielzahl von Inhalten auf YouTube.com von den Nutzern ‚illegal‘ hochgeladen werden, sucht Google derzeit nach Möglichkeiten, die Urheber des Materials auf der Seite an den Werbeumsätzen zu beteiligen, um auf diesem Weg möglichen Schadensersatzklagen wegen Urheberrechtsverletzungen zu vermeiden.⁵⁵

Das starke Umsatzwachstum der Suchmaschinen ist nicht nur eine Folge der geschäftlichen Expansion, sondern vor allem der marktbeherrschenden Stellung auf dem Werbemarkt, die mit einer steigenden Suchmaschinennutzung zusammenfällt. Die Zahl der Suchanfragen stieg in den USA zwischen 2004 und 2005 um 55 Prozent. Allein im Dezember 2005 wurden 5,1 Milliarden Suchanfragen gestellt.⁵⁶ Diese Zahl erklärt die wirtschaftliche Macht von Suchmaschinen, denn 5,1 Milliarden Suchanfragen bedeuten 5,1 Milliarden potenzielle Werbekontakte – Zahlen, die mit anderen Online-Seiten nicht zu erreichen sind und die daher auch die Konzentration der Werbebudgets bei den Suchmaschinen erklären.

⁵⁰ TimeWarner 2006.

⁵¹ Spiegel Online 2006c.

⁵² Ebenda.

⁵³ Google 2006b.

⁵⁴ Vgl. Heise online 2006c.

⁵⁵ Heise online 2006g.

⁵⁶ Nielsen Netratings 2006.

Von größter Bedeutung bei der Suchmaschinenwerbung ist die Vermarktung bezahlter Links, das so genannte Suchmaschinenmarketing. Dabei werden über, neben oder unter einer Ergebnisliste auf Basis der eingegebenen Suchwörter gezielt Anzeigen eingeblendet („Keyword Advertising“), die in ähnlichem Format wie die direkten Suchergebnisse erscheinen. Der Preis dieser „Ad-Links“ für den Werbenden richtet sich nach der Attraktivität der Suchbegriffe und der Klickhäufigkeit.

Der Anteil des Keyword Advertising an der Online-Werbung lag in den USA 2005 bei 41 Prozent bzw. 5,142 Milliarden US-Dollar.⁵⁷ Davon entfielen 49 Prozent auf Google.⁵⁸ Der Gesamtanteil der Suchmaschinen am US-Werbemarkt liegt aber noch höher als 41 Prozent für das Keyword Advertising, da die Suchmaschinen und vor allem Seiten mit Portalcharakter wie Yahoo auch klassische Bannerwerbung betreiben. Auch in Deutschland besitzt die Suchwortvermarktung eine herausragende Bedeutung. 2005 wurden damit 245 Millionen Euro erzielt, was 28 Prozent des Gesamtvolumens der Online-Werbung entsprach.⁵⁹ Suchmaschinen besitzen somit auch eine nicht zu unterschätzende wirtschaftliche Machtposition.

Google ist mit seinem AdWords-Programm Marktführer im Bereich des Keyword Advertisings. Yahoo besitzt mit dem 2003 für 1,6 Milliarden US-Dollar erworbenen Anbieter Overture ebenfalls ein eigenes Vermarktungsinstrument für kontextsensitive Werbung.⁶⁰ Microsoft stellt für das Geschäftsjahr 2006/2007 insgesamt 1,6 Milliarden US-Dollar bereit, um neben der Verbesserung der Suchtechnologie auch ein eigenes System für Suchwortwerbung aufzubauen.⁶¹

Die derzeitige Situation auf dem Suchmaschinenmarkt ist durch ein doppeltes Oligopol gekennzeichnet, das sich einerseits aus der Konzentration auf dem Anbieter- und Nutzermarkt und andererseits auf dem Markt der Suchmaschinenwerbung ergibt. Somit besitzen Suchmaschinen nicht nur die potenzielle Macht, den *Zugang von Inhalten* etwa bestimmter politischer Gruppierungen zu beeinflussen oder zu verhindern, sondern auch bestimmte *Werbekunden* zu diskriminieren.⁶²

Das Keyword Advertising besitzt eine Kehrseite. Denn während Unternehmen ein Interesse daran haben, ihre Anzeigen zu bestimmten Suchbegriffen zu platzieren, wollen die Nutzer vor allem relevante Inhalte finden. Dies führt zu einem Zielkonflikt auf Seiten der Suchmaschinenbetreiber zwischen maximalem Umsatz durch bestmöglichen Kontakt zum Kunden und bestmöglicher Performance auf Nutzerseite, bei der nur treffende Ergebnisse zählen.

Diese Gratwanderung stellt Suchmaschinen vor neue Herausforderungen. Eindeutige Richtlinien für die Platzierung von Werbung fehlen oftmals; seriöse Anzeigenkunden mit ausgefallenen Suchbegriffen werden mitunter abgewiesen.⁶³ Gleichzeitig fehlt die

⁵⁷ IAB 2006.

⁵⁸ eMarketer 2006.

⁵⁹ BVDW 2006a.

⁶⁰ Heise online 2003.

⁶¹ Netzeitung 2006.

⁶² Vgl. Schulz/Held/Laudien 2005, S. 67ff.

⁶³ Beiler/Zenker 2006, S. 14f.

Kontrolle der Vertrauenswürdigkeit eines Anzeigenkunden durch den Suchmaschinenbetreiber, sodass auch unseriöse Werbeanzeigen in den Ergebnislisten auftauchen und die Gefahr, einer manipulierten Website aufzusitzen, in den bezahlten Ergebnissen sogar größer ist als in den normalen Ergebnissen.⁶⁴

Ein weiteres Problem im Bereich der Suchmaschinenökonomie stellt die Anfälligkeit der automatisierten Such- und Rankingalgorithmen für externe Manipulation dar. Um Suchmaschinen herum ist ein kompletter Wirtschaftszweig entstanden, der sich mit der Optimierung von Seiten für bessere Rangpositionen in den Ergebnislisten beschäftigt.⁶⁵

So genannte Search Engine Optimizer nutzen vielfältige Methoden, um Webseiten für bestimmte Suchwörter anzupassen und auf einen besseren Listenplatz im Ranking zu befördern. Dies beeinflusst die Qualität der Ergebnisse von Suchmaschinen, da die Relevanz der so „optimierten“ Webseiten für Suchanfragen oftmals nicht gegeben ist.⁶⁶

Die Grenzen zwischen erlaubtem und illegalem „Tunen“ der Rankingplätze sind dabei fließend. Zwar weisen die Suchmaschinenbetreiber auf die illegalen Praktiken hin, die zu einem Ausschluss führen können, jedoch sind die Kriterien nicht trennscharf. Zudem fehlen Einspruchsmöglichkeiten für die Contentanbieter. Die Intransparenz bezüglich der Ausschlusskriterien kann dazu führen, dass selbst eine Firma wie BMW – wie im Februar 2006 – wegen unerlaubter Optimierungsmethoden kurzzeitig aus dem Google-Index verbannt wird.⁶⁷

Um die im Zusammenhang mit dem Keyword Advertising und der Optimierung von Internetseiten genannten Probleme zu minimieren, vergibt in Deutschland der Bundesverband Digitale Wirtschaft (BVDW) nach genauer Prüfung bestimmter Kriterien Zertifikate an seriöse und transparent arbeitende Suchwortvermarkter und Suchmaschinen-Optimierer.⁶⁸

4 SUCHMASCHINEN UND JOURNALISMUS

Im Forschungsfeld Suchmaschinen und Journalismus stellt sich zum einen die Frage, wie Suchmaschinen von Journalisten als *Rechercheinstrument* eingesetzt werden und inwieweit sich dies auf die Qualität der journalistischen Produkte auswirkt,⁶⁹ und zum anderen, inwieweit *Suchmaschinen selbst journalistische Funktionen übernehmen* und dadurch das Gefüge des bisherigen Massenkommunikationssystems verändern.⁷⁰

Studien zeigen einen wachsenden Einfluss des Internets auf die Arbeitsabläufe und die Wahrnehmung von Journalisten.⁷¹ Zugleich wird der Einfluss des Internets auf die

⁶⁴ Vgl. Edelman 2007.

⁶⁵ Vgl. Van Couvering 2007.

⁶⁶ Vgl. Machill/Neuberger/Schindler 2002.

⁶⁷ Heise online 2006b.

⁶⁸ BVDW 2006b.

⁶⁹ Vgl. Wyss/Keel in Machill/Beiler 2007.

⁷⁰ Der Lehrstuhl für Journalistik II der Universität Leipzig hat 2007 im Auftrag der Landesanstalt für Medien Nordrhein-Westfalen eine Multimethodenstudie zur „Journalistischen Recherche im Internet“, speziell mittels Suchmaschinen, übernommen. Weitere Informationen sind unter www.lehrstuhl-journalistik2.de zu finden.

⁷¹ Vgl. Middleberg/Ross 2006; Keel/Bernet 2005; News aktuell 2002.

journalistische Recherche in der Fachliteratur und in der Praxis kontrovers diskutiert, zumal vor dem Hintergrund einer in Deutschland wenig ausgeprägten Recherchekultur.⁷² In den USA zeigt ein Vergleich von Middleberg und Ross über sieben Jahre, wie rasant das neue Medium in den USA zu einem selbstverständlichen Arbeitsmittel für Journalisten wurde.⁷³ So gingen 1994 erst 17 Prozent der Befragten mindestens täglich online auf Informationssuche, 2000 waren es schon 81 Prozent. Die große Mehrzahl der Befragten zeigte sich zufrieden mit den Veränderungen ihrer Arbeit durch das Internet.

Eine den amerikanischen Studien vergleichbare deutsche Untersuchungsreihe stammt von der dpa-Tochter news aktuell.⁷⁴ Sie zeigt ähnliche Trends auf. Die Onlinenutzung für die redaktionelle Arbeit stieg von 38 Prozent im Jahr 1997 auf 98 Prozent im Jahr 2002. Den größten Teil dieser steigenden Onlinezeit nutzen die deutschen Journalisten für Recherchen. Insbesondere ermöglicht das Internet eine einfache Möglichkeit der Vorrecherche und Zugriff rund um die Uhr. Als mit Abstand wichtigstes Online-Rechercheinstrument sehen 82 Prozent der deutschen Journalisten dabei Suchmaschinen und Webkataloge an.⁷⁵

Keel und Bernet legten zwei repräsentative Befragungen deutschschweizer Print- und Rundfunkjournalisten vor, die einen Vergleich der Jahre 2002 und 2005 ermöglichen.⁷⁶

Der Anteil von Journalisten, die das Internet täglich für ihre Arbeit verwenden, stieg von 80 auf 92 Prozent; zugleich hat sich der Zeitaufwand für Onlinerecherchen deutlich erhöht. Suchmaschinen spielen im Arbeitsalltag der deutschschweizer Journalisten eine zentrale Rolle. Sie werden von 97 Prozent der Befragten als „wichtig“ oder „sehr wichtig“ eingeschätzt. Gleichzeitig lässt sich eine zunehmende Dominanz der Suchmaschine Google feststellen, die von ebenfalls 97 Prozent der Journalisten als „meistgenutzte Suchmaschine“ bezeichnet wird. Die große Abhängigkeit von Suchmaschinen allgemein und von Google im Speziellen wird dabei nur vereinzelt von den Journalisten als Problem erkannt.⁷⁷

Dieser kursorische Überblick zeigt, mit welcher Geschwindigkeit sich die Online-Recherche im Allgemeinen und Suchmaschinen im Speziellen sowohl in den USA, als auch in der Schweiz und in Deutschland in den Redaktionen etabliert haben. Innerhalb weniger Jahre ist das Internet zu einem selbstverständlichen Arbeitsmittel und sind Suchmaschinen zu einem alltäglichen Rechercheinstrument der Journalisten geworden. Neben den Vorteilen der Onlinerecherche, die journalistisches Handeln bei der Vorrecherche unter Zeitdruck sowie durch die zeitlich und räumlich nicht beschränkte Verfügbarkeit von Informationen im Internet erleichtern kann,⁷⁸ birgt die Onlinerecherche Risiken, wenn auf Recherchemethoden außerhalb des Internets weitgehend verzichtet wird.⁷⁹ Die bisherigen Studien deuten darauf hin, dass trotz dieser Entwicklung klassische Recher-

⁷² Vgl. Netzwerk Recherche 2006; Schmitz 2006; Höhne/Ruß-Mohl 2005; Wegner 2005.

⁷³ Middleberg/Ross 2000.

⁷⁴ News aktuell 2000 und 2002.

⁷⁵ News aktuell 2002, S. 5.

⁷⁶ Keel/Bernet 2005.

⁷⁷ Vgl. Keel/Wyss 2006, S. 16-18.

⁷⁸ Vgl. Meier 2002; Lünenbürger-Reidenbach/Petersen/Wagensonner 2000, S. 7; Meier 1997.

⁷⁹ Vgl. Meier 2004, S. 252.

chemethoden offenbar nicht in dem Ausmaß an Bedeutung verloren haben, wie Kritiker befürchtet haben. Das Internet scheint die anderen Recherchequellen somit nicht zu verdrängen, sondern zu ergänzen. Dennoch muss im Folgenden ein Blick auf einige kritische Punkte der Onlinerecherche geworfen werden.

Diese muss sich an dem allgemeinen Selbstverständnis der Journalisten messen lassen, in erster Linie neutral zu berichten bzw. die Realität abzubilden, dem sich 89 bzw. 74 Prozent der deutschen Journalisten verpflichtet fühlen.⁸⁰ Gleichwohl birgt die unreflektierte Nutzung des Internets durch den Journalisten die Gefahr einer verzerrten Berichterstattung über die Wirklichkeit. Das Problem besteht in der Glaubwürdigkeit und Qualität der Quellen, denn diese lassen sich einfach manipulieren und sind keinen professionellen Qualitätsmaßstäben unterworfen,⁸¹ da aufgrund der niedrigen Zugangsbarrieren nahezu jeder im Internet zum Kommunikator werden kann, was zu einer Unmenge von „Informationsmüll“ geführt hat⁸².

Zu einer Verzerrung der Wirklichkeitswahrnehmung kann es bei der Nutzung von Suchmaschinen als journalistischem Recherchewerkzeug auch deshalb kommen, weil die Ranking- und Aktualisierungs-Algorithmen der Suchmaschinen stark selektiv sind. Bestimmte Webseiten – und damit auch die auf diesen Webseiten vertretenen Meinungen und Standpunkte – haben keine Chance, auf wahrnehmbaren Rankingplätzen gelistet zu werden oder überhaupt in den Ergebnislisten von Suchmaschinen aufzutauchen.⁸³ Zudem werden zahlreiche Seiten nur unregelmäßig abgefragt und sind somit nur mit alten Versionen im Index der Suchmaschine präsent.⁸⁴ Hinzu kommt, dass die Ergebnislisten von Suchmaschinen, wie in Abschnitt 3 dargestellt, anfällig für die Manipulation von außen sind.⁸⁵

Verstärkt wird dieses allgemeine Problem von Suchmaschinen durch die Gefahr der Abhängigkeit von einer einzigen Suchmaschine. So warnt Wegner vor einer „Googleisierung der Recherche“, da Google „neuerdings den kleinsten gemeinsamen Nenner für eine Vorrecherche zu definieren“ scheine.⁸⁶ Mitunter wird die Frage gestellt, ob die Folge „eine journalistische Selbstreferenzialität einer ganz neuen Dimension“ sei,⁸⁷ weil im Grunde nur noch bereits veröffentlichte Informationen übernommen werden. Andererseits wird von manchen die These vertreten, dass der Einsatz von Suchmaschinen die journalistischen Mindeststandards gehoben hat, weil auch schlecht mit Recherchekapazitäten ausgestattete, kleine Redaktionen hiermit über eine gute Möglichkeit der Recherche verfügen. So argumentiert Wegner, dass der Einsatz von Suchmaschinen das allgemeine Rechercheniveau sogar verbessert habe.⁸⁸ Weitgehend unbekannt ist jedoch, über welche Kompetenzen Journalisten im Umgang mit Suchmaschinen verfügen.

⁸⁰ Weischenberg/Scholl/Malik 2006, S. 336.

⁸¹ Machill/Beiler 2005, S. 232.

⁸² Machill/Neuberger/Schindler 2002, S. 17.

⁸³ Goldman 2006.

⁸⁴ Vgl. Lewandowski/Wahlig/Meyer-Bautor 2006.

⁸⁵ Gyongyi/Garcia-Molina 2005; Jones 2005; Machill/Neuberger/Schindler 2002.

⁸⁶ Wegner 2002, S. 2.

⁸⁷ Machill/Beiler 2005, S. 232.

⁸⁸ Wegner 2005.

Suchmaschinen wirken nicht nur indirekt als Rechercheinstrument auf den Journalismus ein, sondern übernehmen auch selbst journalistische Funktionen. Sie üben damit eine Doppelrolle im Mediensystem aus. Speziell für die Suche nach aktuellen publizistischen Inhalten haben alle großen Suchmaschinen in den vergangenen Jahren spezielle Suchfunktionen entwickelt.⁸⁹ Google News und der MSN Newsbot betreiben zudem ein automatisiertes Nachrichtenportal, das im Internet gefundene Meldungen automatisch Themen zuordnet und auf einer Seite anordnet, die einem journalistischen Online-Angebot stark ähnelt.

Suchmaschinen stoßen mit diesen speziellen Angeboten also in einen Bereich vor, der bislang dem traditionellen Journalismus vorbehalten war. Dies ist auch vor dem Hintergrund bedeutsam, dass Nachrichten inzwischen der meistgenutzte Inhalt im World Wide Web sind: Nach der Studie „internet facts“ stehen aktuelle Nachrichten zum Weltgeschehen im zweiten Quartal 2006 mit 60 Prozent der Internetnutzer an erster Stelle der zumindest gelegentlich genutzten Inhalte.⁹⁰ Google News und Yahoo! News befinden sich nach Zahlen aus dem Nielsen-Netrating-Panel kontinuierlich in den Top 20 der US-amerikanischen News-Sites. In Deutschland sind Google News und die Yahoo! Nachrichten sogar unter den zehn meistgenutzten Online-Nachrichtenangeboten.⁹¹

Technisch funktionieren Nachrichtensuchmaschinen ähnlich wie normale Web-Suchmaschinen. Crawler durchsuchen Webseiten und indexieren deren Inhalte. Der zentrale Unterschied zu universellen Web-Suchmaschinen ist, dass die Art der Quellen grundsätzlich auf journalistische Angebote eingeschränkt ist.⁹² Die Quellenauswahl ist aber einer der kritischsten Punkte bei der Nachrichtensuche. Denn sie entscheidet darüber, von welchen Angeboten Nachrichten vermittelt werden. Dies kann erhebliche publizistische Auswirkungen haben, denn nicht einbezogene Quellen sind unsichtbar. Dabei ist die Quellenauswahl völlig in das Belieben der Anbieter gestellt: Bei den meisten Nachrichtensuchmaschinen wird weder offengelegt, nach welchen Kriterien die Auswahl getroffen wird, noch welche Angebote durchsucht werden.

Google News ist hierfür ein negatives Beispiel: Zwar können Nutzer und Content-Anbieter die Aufnahme vorschlagen, jedoch weist Google News darauf hin, dass keine Garantie und kein Rechtsanspruch auf die Aufnahme bestünden. Dies widerspricht der Selbstdarstellung, dass Google News „ausschließlich von Computeralgorithmen und völlig ohne das Eingreifen von Menschen erstellt“ werde.⁹³ Zu den Auswahlkriterien heißt es nebulös, dass die Quellen „verschiedene Blickwinkel und redaktionelle Ansätze“ verfolgen. Ihre Auswahl werde „ohne Bezug auf politische Standpunkte oder Ideologien getroffen“.⁹⁴ Eine solche intransparente und intersubjektiv nicht nachvollziehbare Quellenauswahl ist problematisch.

⁸⁹ Machill/Lewandowski/Karzauninkat 2005.

⁹⁰ AGOF 2006, S. 15.

⁹¹ Machill/Beiler 2006, S. 20.

⁹² Vgl. Machill/Lewandowski/Karzauninkat 2005.

⁹³ Google 2003.

⁹⁴ Ebenda.

Bei der Auswahl von Quellen müssen die Betreiber entscheiden, ob auch quasi-journalistische Angebote wie zum Beispiel Weblogs einbezogen werden. Ein Problem hinsichtlich der Abgrenzung journalistischer Angebote kann entstehen, wenn Pressemitteilungen aufgenommen werden. Hier ist Google News in die Kritik geraten, weil die Nachrichtensuchmaschine Pressemitteilungen wie aktuelle Nachrichten behandelte.⁹⁵ Damit verschwimmt die Trennung von redaktionellen Inhalten und PR.

Bei Nachrichtensuchmaschinen besteht zudem das Problem der Konzentration der Meldungen auf wenige Quellen. Dies belegen exemplarische Untersuchungen sowohl für die automatisierten Nachrichtenportale als auch für die Suchfunktion. So stammen nach einer Studie von Machill, Lewandowski und Karzauninkat bei den Altavista-Nachrichten und Paperball 75 Prozent der Meldungen von nur zehn verschiedenen Angeboten, während bei Google News die Spannbreite etwas höher ist (38 Prozent werden von zehn Quellen abgedeckt).⁹⁶ Nach journalistischen Qualitätskriterien sollten die Nachrichtensuchmaschinen um strukturelle und inhaltliche Vielfalt bemüht sein. Es stellt sich die Frage, wie ähnlich die Selektion und das Ranking von Nachrichtensuchmaschinen und die klassische journalistische Auswahl und Vermittlung von Informationen sind.

5 TECHNIK UND QUALITÄT VON SUCHMASCHINEN

Die Marktdominanz Googles im Bereich der Suchmaschinen beruht auf dem technologischen Vorsprung, den der Konzern durch seine 1998 entwickelten Methoden des Rankings von Webseiten nach dem so genannten PageRank gegenüber den Wettbewerbern aufgebaut hat.⁹⁷ Diese Marktführerschaft wird von den Mitbewerbern attackiert. Jahr für Jahr wenden Suchmaschinen einen Großteil ihrer Investitionen für die Verbesserung der Suchtechnologie auf. Bei Google liegt der durchschnittliche Anteil der Weiterentwicklung der Suchtechnologie an den Gesamtinvestitionen bei rund 70 Prozent. Die übrigen 30 Prozent werden in neue Services wie Google Earth und in neue Grundlagentechnologien investiert.⁹⁸ 2005 gab Google 484 Millionen US-Dollar für Forschung und Entwicklung aus⁹⁹, Konkurrent Yahoo 547 Millionen US-Dollar¹⁰⁰.

Diese Zahlen belegen, welche Investitionen notwendig sind, um eine technisch konkurrenzfähige, universelle Suchmaschine aufzubauen und weiterzuentwickeln. Damit bestehen aufgrund einer technisch-finanziellen Sondersituation geringe Markteintrittschancen für neue Suchmaschinenwettbewerber, da hohe Kosten notwendig sind, um den technologischen Vorsprung der Marktführer Google, Yahoo und Microsoft aufzuholen.

Daher ist das in Abschnitt 2 und 3 dargestellte Oligopol der US-Suchmaschinen Google, Yahoo und Microsoft auch unter technologischen Aspekten zu problematisieren. So hat das von dem ehemaligen französischen Staatspräsidenten Chirac angeregte und mit 400 Millionen Euro geförderte französisch-deutsche Projekt Quaero zum Ziel,

⁹⁵ Vgl. z. B. Orłowski 2003.

⁹⁶ Machill/Lewandowski/Karzauninkat 2005, S. 152f.

⁹⁷ Vgl. Brin/Page 1998.

⁹⁸ Beiler/Zenker 2006, S. 12.

⁹⁹ Google 2006a, S. 41.

¹⁰⁰ Yahoo 2006, S. 59.

auch europäischen Firmen eine den drei US-Suchmaschinen ebenbürtige Suchtechnik zu verschaffen.¹⁰¹ Ein vielversprechender Teil des Projekts ist die Suchmaschine Exalead, deren Index bereits über sechs Milliarden Seiten enthält und die fortschrittliche Technologien einsetzt. Nach dem Auseinanderbrechen des Projekts wird die französische Seite mit Exalead weiterhin an einer Konkurrenzsuchmaschine zu Google arbeiten, während in Deutschland mit der Suchmaschine Theseus Grundlagenforschung im Bereich der semantischen Suche vorangetrieben werden soll.¹⁰²

Die Qualität einer Suchmaschine lässt sich nach Lewandowski in vier Dimensionen unterteilen.¹⁰³ Dies sind erstens die *Index-Qualität*, also die Größe und Aktualität des Datenbestandes einer Suchmaschine, zweitens die *Qualität der Treffer*, ob also ein Ergebnis für eine Anfrage relevant ist oder nicht, drittens die *Recherche-Qualität*, welche Funktionen den Nutzer bei seiner Suche unterstützen, und viertens die *Usability*, ob die Nutzung der Suchmaschine für den Nutzer einfach und verständlich ist und ob die Suchergebnisse in geeigneter Form aufbereitet werden. Diese vier Qualitätsdimensionen führen zu technischen Entwicklungen in unterschiedlichen Bereichen. Denn die Technologie der etablierten, großen Suchmaschinen ist längst noch nicht ausgereift.

Durch neue, bahnbrechende Technologien ist es somit möglich, dass neue Anbieter sich einen signifikanten Marktanteil sichern können. Es gibt viele Bereiche, in denen die Technologie der Suchmaschinen ausbaufähig ist. Hierzu gehört die Erfassung des kompletten Internets. Marktführer Google erfasste Ende 2006 über acht Milliarden Webseiten, erkennt damit aber trotzdem nur einen Bruchteil des auf über 550 Milliarden Einzelseiten geschätzten Internets.¹⁰⁴ Suchmaschinen können somit immer nur einen Ausschnitt des gesamten Netzes durchsuchen. Schwerwiegender als die unvollständige Indexierung des gesamten Netzes ist die Tatsache, dass es auf technischem Weg bisher nicht möglich ist, das Problem der Relevanz von Suchanfragen in den Griff zu bekommen: „In other words, [...] while we seem to be mastering the coverage issue, we continue to struggle with the issue of what precisely to extract from the enormous bulk of possibilities for a given search“.¹⁰⁵

Bis heute existieren zum Beispiel keine einheitlichen Berechnungsmethoden für das Ranking einer Suchmaschine, die sicherstellen, dass bei gleichen Suchanfragen auch gleiche Ergebnislisten dargestellt werden. Selbst wenn sie dieselben Internetseiten indexieren, geben Suchmaschinen oftmals unterschiedliche Ergebnisse aus. Dies hängt mit den unzureichenden Möglichkeiten der Indexierung von Internetinhalten zusammen. Die Indexierung beruht vor allem auf statistischen und nicht auf semantischen Verfahren, zieht beispielsweise die Häufigkeit von Stichwörtern in die Berechnung der Wichtigkeit einer Seite ein, aber nicht die inhaltlichen Zusammenhänge, in denen diese Stichwörter auftauchen. Suchmaschinen nutzen also mathematische Hilfskonstrukte, die von Suchmaschine zu Suchmaschine variieren, um auf die Relevanz des Inhalts zu schließen.

¹⁰¹ Heise online 2006e; Beiler/Machill 2007.

¹⁰² Stöcker/Darnbeck 2006.

¹⁰³ Lewandowski 2007.

¹⁰⁴ Bergman 2000.

¹⁰⁵ Introna/Nissenbaum 2000, S. 15.

Für den Nutzer ist dieser Mangel aufgrund des Kommunikationsverhaltens der Suchmaschinenbetreiber nicht sofort erkennbar, verdeckt die fehlende Transparenz bei der Selbstdarstellung doch die Problemfelder, die die Qualität der Suchergebnisse negativ beeinflussen. So werden die Variablen, die für die Zuordnung von Relevanz intern zugrunde gelegt werden, nicht öffentlich gemacht. Zwar verweisen Suchmaschinen darauf, welche Faktoren allgemein einen Einfluss auf die Relevanz einer Webseite haben, aber nicht, wie diese beim Ranking genau gewichtet werden. Oftmals ist zudem nur die Hälfte aller Suchergebnisse für eine Suchanfrage relevant¹⁰⁶, und unterschiedliche Webseiten besitzen unterschiedliche Chancen, überhaupt indexiert zu werden¹⁰⁷. Suchmaschinen verzerren somit die Wirklichkeit des Internets,¹⁰⁸ ohne dass dies von außen für den Nutzer erkennbar wäre. Eine Evaluation der Ergebnisse durch den Nutzer ist somit schwer möglich. Als Hauptgrund für die ‚Geheimniskrämerei‘ nennen Suchmaschinen den Schutz von Geschäftsgeheimnissen und den Schutz vor Manipulationen.

Das Beispiel der *Google-Bomb*¹⁰⁹ „miserable failure“, bei der Hunderte von Webseiten Links unter diesem Namen zur Biografie von George W. Bush auf whitehouse.org anlegten, worauf unter diesem Suchbegriff die Bush'sche Biografie auf dem ersten Listenplatz auftauchte,¹¹⁰ zeigt noch auf humorvolle Weise, wie anfällig die Rankingmechanismen auf Manipulation von außen reagieren. Doch diese Manipulation, die aus den unterschiedlichsten Gründen vorgenommen wird, führt zu einem massiven Verlust an Qualität in den Ergebnislisten der Suchmaschinen. Seiten mit jugendgefährdenden Inhalten wie Pornografie finden sich ebenso auf den vorderen Plätzen der Suchergebnisse¹¹¹ wie Seiten mit krimineller Motivation¹¹². Die Anfälligkeit für unterschiedliche Methoden der Optimierung respektive Manipulation von Webseiten – die Grenzen sind oftmals fließend – ist bei den einzelnen Suchmaschinenbetreibern unterschiedlich stark ausgeprägt.¹¹³ Um manipulierte und unerwünschte Seiten aus ihrem Index zu entfernen, greifen die Suchmaschinenbetreiber primär auf automatische Spam-Filter zurück, die permanent an neue Methoden der Manipulation angepasst werden müssen.¹¹⁴

Eine Verbesserung der Qualität der Ergebnisse von Suchmaschinen soll durch die semantische Analyse von Internetinhalten erreicht werden. Bereits heute werden Technologien wie das „Clustering“ in Suchmaschinen angewandt, das Stichworte zu Gruppen zusammenfasst. Beim Clustering werden alle Ergebnisseiten einer Suchanfrage in Untergruppen unterteilt, die in Bezug auf bestimmte Stichworte Gemeinsamkeiten aufweisen;

¹⁰⁶ Véronis 2006; Griesbaum 2004.

¹⁰⁷ Lawrence/Giles 1999.

¹⁰⁸ Goldman 2006; Vaughan/Thelwall 2004.

¹⁰⁹ Mit „Google-Bomb“ wird die bewusste Manipulation von Google bezeichnet, bei der der Page-Rank-Mechanismus ausgenutzt wird, um eine Seite im Ranking auf einen vorderen Listenplatz zu befördern. Viele externe Seiten legen Links unter einem oder mehreren ganz bestimmten Stichwörtern auf die im Ranking zu pushende Seite an, woraufhin diese für die benutzten Stichwörter einen sehr hohen Rankingplatz erhält.

¹¹⁰ Rötzer 2005.

¹¹¹ Véronis 2006; Machill/Neuberger/Schindler 2002.

¹¹² Edelman/Rosenbaum 2006.

¹¹³ Vgl. Machill/Neuberger/Schweiger/Wirth 2003, S. 125ff.

¹¹⁴ Vgl. Adler 2005, S. 75ff.

so lassen sich beispielsweise zum Suchbegriff „Tiger“ die drei Unterkategorien „Raubtier“, „Mac-Betriebssystem“ sowie „Kampfpanzer WWII“ unterscheiden. Der Nutzer erhält so die Möglichkeit, eine dieser Unterkategorien auszuwählen, um seine Suche auf die entsprechende Unterkategorie und die zugehörigen Seiten einzuschränken. Durch das Clustering ist es dem Nutzer somit möglich, seine Suchanfrage während der Suche noch weiter zu spezifizieren.¹¹⁵

Ebenso wichtig für die Qualität einer Suchmaschine ist die Präsentation der Ergebnisse. Bisher folgt die Darstellung von Suchergebnissen bei den Markführern einem hierarchischen Muster. Dabei werden die Ergebnisse einer Suchanfrage in einer Rangfolge angeordnet, die auf den von der Suchmaschine zugrunde gelegten Relevanzkriterien basiert. Als Vorschau dient in der Regel nur ein Textausschnitt, der den Suchbegriff in dem ihn umgebenden Kontext zeigt. Doch es gibt zahlreiche Ansätze, die versuchen, diese Unzulänglichkeiten der Ergebnisrepräsentation zu umgehen. Exalead (www.exalead.de) zeigt zum Beispiel zu jedem Treffer eine Miniaturvorschau der Seite. Seekport bietet die Möglichkeit, die Treffer innerhalb der Suchanfrage in einem Vorschauenfenster anzuzeigen (www.seekport.de).

Eine Mischung zwischen Suchwort-Clustering und grafischer Darstellung der Verlinkung einer Seite bietet Kartoo. Die Suchmaschine ordnet Suchergebnisse nicht in einer Liste, sondern generiert „Landkarten“ mit Ergebnissen, auf denen die Verbindungen zwischen den einzelnen Seiten und den Suchbegriffen dargestellt werden. Die Relevanz einzelner Seiten wird durch die Größe von Schrift und „Landflächen“ visualisiert. Dies eröffnet dem Suchenden neue Freiräume, da die Ergebniskarte einer Mindmap ähnelt und nicht der strengen Hierarchie von *Ergebnislisten* folgt. Durch die grafische Visualisierung werden die Suchergebnisse somit strukturiert und geordnet, wodurch das Problem des „Information Overload“ reduziert wird.¹¹⁶

Allgemein bleibt festzuhalten, dass das Benutzerinterface, mit dem der Zugriff auf den Index der Suchmaschine erfolgt, einfach zu bedienen sein muss, so wie es Google mit seiner minimalistischen Startseite vormacht. Der Gestaltung derartiger Schnittstellen zwischen Mensch und Maschine und der damit einhergehenden Nutzbarkeit kommt gerade im Internet eine wichtige Rolle zu.¹¹⁷ Denn die Nutzbarkeit einer Suchmaschine hat einen wesentlichen Einfluss auf den Erfolg, mit dem ein Nutzer eine Suchanfrage ausführen kann.¹¹⁸

Um die Qualität von Suchergebnissen künftig zu verbessern, müssen Inhalte präziser erkannt und eingeordnet werden können. Eine semantische Analyse von Internetinhalten ist derzeit noch nicht möglich. Die HTML-Sprache bietet kaum Möglichkeiten, um einer Suchmaschine verständlich zu machen, welcher Teil einer Webseite inhaltlich wichtig ist. Suchmaschinen nutzen für die Ermittlung der Relevanz bislang unter anderem Teile von HTML, etwa die Definitionen von Überschriften. Dabei ist HTML ursprünglich als Seitenbeschreibungssprache entwickelt worden, die der Gliederung eines Textes und

¹¹⁵ Vgl. Jiang/Joshi/Krishnapuram/Yi 2000, S. 1f.

¹¹⁶ Albertoni/Bertone/De Martino 2004, S. 4.

¹¹⁷ Bruno/Tam/Thom 2005.

¹¹⁸ Beiler 2005; Aurelio/Mourant 2001.

der Darstellung auf dem Bildschirm dienen soll, nicht aber der Zuordnung der Relevanz von Inhalten. Denn ob eine HTML-Überschriften-Definition, beispielsweise auf einer Nachrichtenseite, zum eigentlichen Artikel gehört oder aber Teil der Navigationsstruktur ist, kann eine Suchmaschine nicht erkennen. Der HTML-Nachfolger XML („Extended Markup Language“) soll hier Abhilfe schaffen. XML ermöglicht die Kennzeichnung semantischer Kategorien in Bezug auf die Inhalte von Webseiten.¹¹⁹ Ziel der Entwicklung derartiger Technologien ist das *Semantic Web*.

Das „Semantic Web“¹²⁰ soll die automatische Zuordnung von Inhalten zu Kategorien wie Name, Wohnort, Telefon und Ähnlichem ermöglichen und so Software in die Lage versetzen, die Inhalte des Netzes gezielt und sinnvoll automatisch zu erschließen. Suchmaschinen könnten in dieser Bewegung eine große Rolle spielen und die Umsetzung des Semantic Web beschleunigen, indem beispielsweise Kategorien der Informationsbereitstellung definiert werden, die die automatische Klassifizierung mehrsprachiger Inhalte ermöglichen. Suchmaschinen könnten durch diese technische Weiterentwicklung und in einer solchen semantischen Umgebung ihre Qualität als Gatekeeper steigern. Einen ersten Schritt gingen die drei Marktführer Google, Yahoo und Microsoft im November 2006, als sie einen gemeinsamen Standard für Sitemaps, basierend auf XML, vorstellten.¹²¹ Das Semantic Web kann somit die Qualität von Suchmaschinen verbessern, führt aber nicht automatisch zu intelligent agierenden Suchmaschinen.¹²² Das Semantic Web würde aber die Möglichkeiten der Klassifizierung von Webseiten verbessern und somit die Entwicklung intelligenter Suchmaschinen erleichtern¹²³ sowie die Entwicklung von semantischen Suchmaschinen fördern.

6 NUTZERVERHALTEN UND -KOMPETENZ

In der Suchmaschinenforschung dürfen die Nutzer bzw. das Nutzerverhalten nicht vernachlässigt werden. Denn die Suche mittels Suchmaschinen ist nur als „transaktionaler Koselektionprozess zwischen Nutzer und Suchmaschine“ zu verstehen,¹²⁴ das heißt sie besteht in einem Zusammenwirken von Mensch und Maschine. Wenn man versteht, aus welchen Motiven, mit welcher Fertigkeit und mit welchem Erfolg Nutzer Suchmaschinen einsetzen, kann dies helfen, die Handhabbarkeit und Qualität von Suchmaschinen sowie die Nutzerkompetenz zu verbessern.

Hierzu tragen Modelle der Suchmaschinennutzung bei, die in verschiedenen wissenschaftlichen Disziplinen entwickelt worden sind.¹²⁵ Solche Modelle helfen, den Suchprozess zu verstehen und die Forschung zu bestimmten Aspekten des Suchprozesses zu systematisieren. Ein umfassendes empirisches Modell des Selektionsprozesses mittels Suchmaschine hat unter anderem Hölischer entwickelt.¹²⁶ Am Anfang eines Suchpro-

¹¹⁹ Green 2000, S. 132f.

¹²⁰ Vgl. Berners-Lee 1998.

¹²¹ Heise online 2006d.

¹²² Vgl. Berners-Lee 1998.

¹²³ Drost 2005.

¹²⁴ Beiler 2005, S. 186; vgl. auch Wirth/Schweiger 1999, S. 66f.

¹²⁵ Vgl. Beiler 2005, S. 167.

¹²⁶ Hölischer 2002, S. 101-112; vgl. auch Marchionini 1995; Shneiderman/Byrd/Croft 1997.

zesses steht dabei ein Informationsbedürfnis (Schritt 0). Falls der Suchende keine entsprechende Webseite kennt, wird er eine Suchmaschine aufrufen (Schritt 1). Es schließen sich das Generieren von potenziellen Suchbegriffen und die Spezifizierung der Suchanfrage an (Schritt 2 und 3). In Schritt 4 wird die Suchanfrage abgeschickt und damit die Ergebnisliste aufgerufen. Es folgt die Evaluation der Suchergebnisse in der Ergebnisliste (Schritt 5). Wird ein Suchergebnis angeklickt (Schritt 6), betrachtet und bewertet der Suchende dieses Dokument (Schritt 7). Enthält es das Gesuchte, endet die Suche (Schritt 10). Andernfalls browsst der Nutzer ausgehend von diesem Dokument weiter (Schritte 8 und 9) oder er kehrt zur Suchmaschine zurück. In diesem Fall muss er entscheiden, ob er neue Dokumente aus der Ergebnisliste aufruft, sich weitere Seiten der Ergebnisliste zeigen lässt oder die Suchanfrage überarbeitet. Der Suchprozess mittels Suchmaschine lässt sich insgesamt als iteratives Vorgehen beschreiben.¹²⁷

Neben solchen beschreibenden Modellen liegen hauptsächlich quantitative Daten zum Nutzerverhalten vor. So wird regelmäßig erhoben, wie viele Nutzer Suchmaschinen einsetzen. Hier kommt die ARD-ZDF-Online-Studie 2006 mittels einer repräsentativen Telefonbefragung auf einen Anteil von drei Vierteln, die mindestens einmal pro Woche auf Suchmaschinen zurückgreifen.¹²⁸ Die ebenfalls repräsentative Studie „internet facts“ der Arbeitsgemeinschaft Online-Forschung weist für das zweite Quartal einen weitesten Nutzerkreis (mindestens gelegentliche Nutzung in den letzten drei Monaten) von 86 Prozent für Suchmaschinen aus.¹²⁹ Auch die Nutzungsreichweiten und -anteile der einzelnen Suchmaschine werden kontinuierlich durch die Panels von Nielsen Netratings und comScore Media Metrix erfasst.¹³⁰

Außer diesen Reichweiten und Suchanteilen gibt es nur wenige Zahlen, die qualitative Aspekte der Suchmaschinennutzung betrachten und auf einer Mikroebene das Verhalten der Nutzer untersuchen. Die wenigen vorliegenden Ergebnisse deuten darauf hin, dass die Nutzer im Umgang mit Suchmaschinen wenig kompetent sind. Der Umgang ist zudem durch Kritiklosigkeit und fehlendes Problembewusstsein geprägt.

So vertrauen 49 Prozent der Suchmaschinennutzer in Deutschland immer und ausschließlich einer einzigen Suchmaschine. Aber auch Nutzer, die neben ihrer Haupt- noch eine Zweitsuchmaschine haben, verwenden diese nur in drei Prozent der Fälle „sehr oft“ und in weiteren acht Prozent der Fälle „oft“, während es bei der Hauptsuchmaschine 77 Prozent sind.¹³¹ Damit lässt sich festhalten, dass etwa die Hälfte der Internetnutzer nie die Suchmaschine wechselt und der gleichberechtigte Einsatz zweier oder mehr Suchmaschinen kaum vorkommt. Auch in den USA setzen 44 Prozent der Nutzer nur auf eine einzige Suchmaschine.¹³² Damit begeben sich die Nutzer in eine Abhängigkeit von den Selektions- und Rankingkriterien einer Suchmaschine. Da bestimmte Suchmaschinen nur

¹²⁷ Vgl. auch den "TOTE-Prozess" bei Miller/Galanter/Pribram 1973; Weber/Groner 1999.

¹²⁸ Van Eimeren/Frees 2006, S. 406.

¹²⁹ AGOF 2006, S. 15.

¹³⁰ Vgl. die einleitend genannten Daten; allgemein dazu Sullivan 2006a.

¹³¹ Machill/Neuberger/Schweiger/Wirth 2003, S. 149.

¹³² Fallows 2005.

bestimmte Teile des gesamten Internets erfassen,¹³³ ergibt sich für Nutzer das Problem, dass sie eine Vielzahl potenziell relevanter Ergebnisse ausschließen.

Schaut man sich die Suchstrategien und -prozesse der Nutzer im Detail an, so bestätigt sich die geringe Nutzungskompetenz. Bei der Formulierung der Suchanfrage verzichten die meisten Nutzer darauf, von den Möglichkeiten der Einschränkung und Präzisierung Gebrauch zu machen. So kennen nur 49 Prozent der Nutzer Suchoperatoren, nur 20 Prozent verwenden sie öfter.¹³⁴ Die erweiterte oder Profisuche kennen zwar immerhin 59 Prozent der Nutzer, allerdings findet sie öfter nur bei 14 Prozent Anwendung.¹³⁵ Die – teilweise schwer zu findenden und komplizierten – Hilfeseiten der Suchmaschinen kennen nur 49 Prozent der Nutzer.¹³⁶ Diese Ergebnisse werden durch andere Studien, etwa Auswertungen von Logfiles, gestützt, wonach der Großteil der Suchanfragen aus Einwortsuchen besteht.¹³⁷

Bei der Evaluation der Suchergebnisse fällt auf, dass ein Großteil der Nutzer nur die erste Seite der Ergebnisliste anschaut und lediglich wenige Treffer aufruft. Logfile-Analysen zeigen, dass durchschnittlich etwa 1,7 Seiten der Ergebnisliste à zehn Suchergebnisse angesehen und maximal zehn Suchergebnisse angeklickt werden.¹³⁸ Auch die konkrete Auswahl der Suchergebnisse ist bei über der Hälfte der Klicks von Nutzern nicht reflektiert.¹³⁹ Allerdings liefert Beiler Hinweise darauf, dass intuitive und heuristische Entscheidungen nicht nur mit einer geringen Kompetenz erklärt werden können, sondern diese auch auf einer höheren Ebene rational und ökonomisch sinnvoll erscheinen, weil eine präzisere Auswahl nicht notwendig oder möglich ist, wenn die Form der Ergebnispräsentation von Suchmaschinen für eine reflektierte Auswahl ungenügend ist.¹⁴⁰

Trotz der größtenteils geringen Kompetenz der Nutzer fühlen sich 92 Prozent der US-Nutzer sicher im Umgang mit Suchmaschinen¹⁴¹ und bezeichnen sich zwei Drittel der deutschen Nutzer als Fortgeschrittene oder Experten¹⁴². Die Selbsteinschätzung entspricht nicht der tatsächlichen Kompetenz der Nutzer. Damit sind Suchmaschinenbetreiber, Bildungsinstitutionen und Medienpolitik aufgerufen, Schritte zur Verbesserung der Nutzerkompetenz zu unternehmen. So hat die Landesanstalt für Medien Nordrhein-Westfalen (LfM) „12 goldene Suchmaschinen-Regeln“ veröffentlicht.¹⁴³ Solche Maßnahmen können auch das Problembewusstsein der Nutzer schärfen. Denn es lassen sich verschiedene Aspekte für eine kritische Auseinandersetzung mit Suchmaschinen ausmachen.

Die Online-Sicherheit in den Ergebnislisten der Suchmaschinen ist in mehrfacher Hinsicht nicht gegeben. So stellt Schindler immer noch gravierende Mängel aus Sicht des Jugendschutzes in den Ergebnislisten fest.¹⁴⁴ Pornografie, Gewaltdarstellung und

¹³³ Vaughan/Thelwall 2004; Lawrence/Giles 1999.

¹³⁴ Machill/Neuberger/Schweiger/Wirth 2003, S. 167.

¹³⁵ Ebenda, S. 168f.

¹³⁶ Ebenda, S. 168.

¹³⁷ Vgl. den Überblick bei Jansen/Pooch 2001.

¹³⁸ Jansen/Pooch 2001.

¹³⁹ Vgl. Machill/Neuberger/Schweiger/Wirth 2003, S. 272f.

¹⁴⁰ Beiler 2005, S. 185f.

¹⁴¹ Fallows 2005.

¹⁴² Machill/Neuberger/Schweiger/Wirth 2003, S. 138.

¹⁴³ Machill 2005.

¹⁴⁴ Schindler 2005.

Seiten mit kriminellem Hintergrund finden sich auch bei vergleichsweise harmlosen Suchanfragen in den Ergebnislisten der Suchmaschinen.¹⁴⁵ Gerade bei populären Suchbegriffen wie „Skype“ besteht die Gefahr, dass Seiten mit Trojanern, Viren und Dialern auf den vorderen Plätzen der Trefferlisten auftauchen.¹⁴⁶ Surft ein wenig kompetenter Nutzer – und gerade bei Kindern ist die Kompetenz im Umgang mit Suchmaschinen und bei der Evaluation von Ergebnisseiten oftmals nicht stark ausgeprägt – eine solche Seite an, können zum Beispiel Trojaner persönliche Daten wie Kontoverbindungen ausspionieren oder aber den Rechner des Nutzers zum Versand von Spam-Mails missbrauchen. Viren können bei fehlendem oder nicht aktuellem Virenschutz zu Datenverlust führen. Dialer, die oftmals als „kostenloses Zugangsprogramm“ getarnt werden, verursachen für den Nutzer erhebliche Mehrkosten.¹⁴⁷

Filtersoftware als Schutz vor derartigen Inhalten ist nur bedingt geeignet. Zwar reduzieren die automatischen Filter der Suchmaschinen die Zahl von auffälligen Ergebnissen, können diese aber nicht komplett verhindern.¹⁴⁸ Separate Filtersoftware oder aber Webseiten-Einstufungssysteme sind zwar im Bereich von jugendschutzrechtlich bedenklichen Inhalten nützlich,¹⁴⁹ schützen aber nicht vor Seiten mit Trojanern, Viren oder Dialern, sofern diese von unbedenklichen Inhalten, zum Beispiel trivialen Informationen über Popstars, umgeben sind.

Die Verwendung von automatischen Filtersystemen auf Seiten der Suchmaschinenbetreiber hat auch eine Kehrseite. Denn während einerseits der Schutz von Kindern und Jugendlichen vor gefährdenden Webseiten ein großes öffentliches Interesse darstellt, steht auf der anderen Seite das Recht auf freien Informationszugang, das durch automatische Filtersysteme eingeschränkt wird. Dies spiegelt das Problem der fehlenden Möglichkeit einer zuverlässigen Altersverifizierung im kostenlosen Internet wider, denn die fehlende Möglichkeit der Alterskontrolle führt dazu, dass auch Erwachsene von für Kinder und Jugendliche gesperrten Seiten ausgeschlossen werden. Auch die unterschiedliche Gesetzgebung beeinflusst, was gefiltert wird und was nicht. Während beispielsweise in den USA Internetseiten mit rechtsradikalen Inhalten unter die Rede- und Meinungsfreiheit fallen, müssen derartige Seiten in Deutschland nach Aufforderung aus den Indizes entfernt werden. Eine bessere *Kontrolle der Kontrolle* wäre hier anzustreben.¹⁵⁰

Um die Relevanz von Ergebnissen bei Suchanfragen zu erhöhen, haben sich mittlerweile zahlreiche Formen von Spezialsuchen etabliert, die nicht die Gesamtheit des Netzes durchsuchen, sondern diese Gesamtheit unter bestimmten Aspekten einschränken. Die wohl bekannteste Form einer solchen Spezialsuche ist die Bildersuche, die mittlerweile von fast jeder Suchmaschine angeboten wird.

¹⁴⁵ Vgl. auch Machill/Neuberger/Schindler 2002.

¹⁴⁶ Vgl. Edelmann 2007.

¹⁴⁷ Machill 2005.

¹⁴⁸ Schindler 2005.

¹⁴⁹ Vgl. auch Frau-Meigs 2007.

¹⁵⁰ Vgl. ebenda.

Da rund 70 Prozent des Internets in multimedialen Inhalten, Grafiken, Flash-Animationen, Musik und Videos bestehen,¹⁵¹ sind derartige Spezialsuchen zum Auffinden bestimmter Inhalte sinnvoll. Doch da Suchmaschinen derartige Inhalte nur aufgrund des sie umgebenden Textes auf der Webseite einordnen können, ein automatisches Erkennen des Inhaltes eines Bildes nicht möglich ist, verlangen Spezialsuchen vom Nutzer eine erhöhte Nutzungskompetenz und auf Seiten der Suchmaschinen verstärkte Qualitätsbemühungen, um unerwünschte Inhalte zu entfernen.

Weitere Spezialsuchen, die von Suchmaschinen angeboten werden, umfassen beispielsweise den Bereich der Produktsuche, der lokalen Suche (in Webseiten aus einer geografisch eingrenzbaaren Region) sowie die Recherche nach wissenschaftlichen Publikationen, wie sie Google Scholar anbietet.¹⁵² Auch bei diesen Spezialsuchen gibt es Probleme. So garantiert dem Nutzer die bloße Listung eines Online-Shops in einer Produktsuchmaschine nicht dessen Vertrauenswürdigkeit, sodass der Nutzer diese über andere Quellen verifizieren muss.¹⁵³

Inwiefern Spezialsuchen bzw. Suchmaschinen allgemein bestimmte Offline-Services wie Telefonbücher oder Bibliotheken ablösen oder aber ergänzen werden, bleibt abzuwarten. Ein Substitutionsdruck zeichnet sich vor allem bei den Medien ab, die eine funktionale Ähnlichkeit zu Suchmaschinen aufweisen, wie dies bei Telefonbüchern der Fall ist.¹⁵⁴

Suchmaschinen sind bereits ein unersetzlicher Teil des Internets. In einer fortschreitend vernetzten Gesellschaft, wie sie zumindest für die Industrienationen festgestellt werden kann, ist es daher notwendig, dass der richtige Umgang mit diesen Hilfsmitteln frühzeitig erlernt wird, dass das Problembewusstsein für die Vorzüge und Schwächen von Internetsuchmaschinen geschärft wird und dass vor allem ein kritischer Umgang mit den unterschiedlichen Suchmaschinen vermittelt wird. Um diese notwendige Medienkompetenz zu fördern, sind Medien- und Bildungspolitikern ebenso gefordert wie die Betreiber der Suchmaschinen sowie die Gesellschaft als Ganze.

¹⁵¹ Green 2000.

¹⁵² Vgl. Machill/Beiler/Neumann 2007.

¹⁵³ Vgl. Machill 2005.

¹⁵⁴ Vgl. Kink/Hess 2007.

7 LITERATUR

Albertoni/Bertone/De Martino 2004

Albertoni, R./Bertone, A. /De Martino, M.: *Semantic Web and Information Visualization*. URL:<http://semanticweb.deit.univpm.it/swap2004/cameraready/albertoni.pdf> [Stand: 2004].

Adler 2005

Adler, T.: „Suchmaschinen-Spamming: Herausforderung für Trefferqualität und Jugendschutz. Ein Insider-Bericht.“ In: Machill, M./Schneider, N. (Hrsg.): *Suchmaschinen: Herausforderung für die Medienpolitik*. Berlin: Vistas Verlag, 2005, S. 71-81.

AGOF 2006

AGOF – Arbeitsgemeinschaft Online-Forschung: *Berichtsband – Teil 1 zur internet facts 2006-II*. URL: <http://www.agof.de/index.download.6a25486a7522c4eb42ab512fbd8e7f4e.pdf> [Stand: 2006].

Aurelio/Mourant 2001

Aurelio, D. N./Mourant, R. R.: „Effect of Web Search Engine Design on User Performance and Preference.“ In: *Proceedings Ninth International Conference on Human-Computer Interaction*. (Tagung) URL: <http://www1.coe.neu.edu/~mourant/pubs/Usab2001.pdf> [Stand: 2001].

Beiler 2005

Beiler, M.: „Selektionsverhalten in den Ergebnislisten von Suchmaschinen. Modellentwicklung und empirische Überprüfung am Beispiel von Google.“ In: Machill, M./Schneider, N. (Hrsg.): *Suchmaschinen: Herausforderung für die Medienpolitik*. Berlin: Vistas Verlag, 2005, S. 165-189.

Beiler/Machill 2007

Beiler, M./Machill, M.: „Suchmaschinen im Internet als Governanceproblem.“ In: Donges, P. (Hrsg.): *Medienpolitik und Media Governance: Was bringt die Governance-Perspektive für die Analyse von Medienpolitik?* Köln: Herbert von Halem Verlag, 2007 S. 250-265.

Beiler/Zenker 2006

Beiler, M./Zenker, M.: *Die wachsende Macht von Suchmaschinen im Internet: Auswirkungen auf User, Medienpolitik und Medienbusiness. Wissenschaftlicher Workshop und Konferenz, 26./27. Juni 2006, Berlin. Zusammenfassung*. URL: http://www.uni-leipzig.de/journalistik/suma/pdf/Zusammenfassung_Suma-Tagung.pdf [Stand: 2006].

Bergman 2000

Bergman, M. K.: *The Deep Web. Surfacing hidden value. White Paper. Bright Planet.* URL: <http://www.brightplanet.com/deepcontent/index.asp> [Stand: 2000].

Berners-Lee 1998

Berners-Lee, T.: *Semantic Web Road Map.* URL: <http://www.w3.org/DesignIssues/Semantic.html> [Stand: 1998].

Brin/Page 1998

Brin, S./Page, L.: *The anatomy of a large-scale hypertextual Web search engine.* URL: <http://www.public.asu.edu/~ychen127/cse591f05/anatomy.pdf> [Stand: 1998].

Bruce Clay 2006

Bruce Clay: *Search Engine Relationship Chart.* URL: <http://www.bruceclay.com/serc.htm> [Stand: 2006].

Bruno/Tam/Thom 2005

Bruno, V./Tam, A./Thom, J.: *Characteristics of web applications that affect usability: a review.* URL: http://portal.acm.org/ft_gateway.cfm?id=1108445&type=pdf&coll=GUIDE&dl=GUIDE&CFID=54814&CFTOKEN=28115244 [Stand: 2005].

BVDW 2006a

BVDW: *885 Millionen Euro für Online-Werbung 2005.* URL: http://www.bvdw.org/shared/data/pressclipping/2300_001_pm_ovk_werbestatistik_060210_bh3.pdf [Stand: 2006].

BVDW 2006b

BVDW: *13 Zertifikate im Bereich Suchmaschinen-Marketing vergeben. Kompetenznachweis für Suchwortwerbung und Suchmaschinen-Optimierung.* URL: http://www.bvdw.org/uploads/media/2300_001_pm_seosem-zertifikat_060830_final2.pdf [Stand: 2006].

CNNIC 2005

CNNIC: *China Online Search Market. Survey Report 2005.* URL: <http://www.cnnic.org.cn/download/2005/2005083101.pdf> [Stand: 2005].

Crowell 2005

Crowell, G.: *Copyrights, Trademarks and Search Engines.* URL: <http://searchenginewatch.com/showPage.html?page=3557706> [Stand: 2005].

Drost 2005

Drost, I.: *Developing Intelligent Search Engines*. URL: http://events.ccc.de/congress/2005/fahrplan/attachments/532-Paper_DevelopingIntelligentSearchEngines.pdf [Stand: 2005].

Edelman 2007

Edelman, B.: „Assessing and Improving the Safety of Internet Search Engines.“ In: Machill, M./Beiler, M. (Hrsg.): *Die Macht der Suchmaschinen/The Power of Search Engines*. Köln: Herbert von Halem Verlag, 2007, S. 259-277.

Edelman/Rosenbaum 2006

Edelman, B./Rosenbaum, H.: *The Safety of Internet Search Engines*. URL: http://www.siteadvisor.com/studies/search_safety_may2006.html [Stand: 2006].

eMarketer 2006

eMarketer: *Search Marketing: Players and Problems*. URL: http://www.emarketer.com/Report.aspx?search_marketing_players_apr06 [Stand: 2006].

Fallows 2005

Fallows, D.: *Search Engine Users. Internet searchers are confident, satisfied and trusting – but they are also unaware and naïve. Pew Internet & American Life Project*. URL: http://www.pewinternet.org/pdfs/PIP_Searchengine_users.pdf [Stand: 2005].

Frau-Meigs 2007

Frau-Meigs, D.: „Minding the Gatekeepers: Search Engines for Young People, and the Regulatory Riddle of Harmful Content – an Environmental Cognition Perspective.“ In: Machill, M./Beiler, M. (Hrsg.): *Die Macht der Suchmaschinen/The Power of Search Engines*. Köln: Herbert von Halem Verlag, 2007, S. 308-326.

FSM 2004a

FSM: *Verhaltenssubkodex für Suchmaschinenanbieter*. URL: http://www.fsm.de/inhalt.doc/Verhaltenssubkodex_SuMa.pdf [Stand: 2004].

FSM 2004b

FSM: *Verfahrensordnung für Verfahren bei Beschwerden gegen den Verhaltenssubkodex der Suchmaschinenanbieter*. URL: http://www.fsm.de/inhalt.doc/Verfahrensordnung_SuMa.pdf [Stand: 2004].

FSM 2005

FSM: *Suchmaschinebetreiber gründen Freiwillige Selbstkontrolle unter dem Dach der FSM*.
 URL: http://www.fsm.de/inhalt.doc/PM_2005-02-24_Suchmaschinen.pdf
 [Stand: 2005].

Gasser/Thurman 2007

Gasser, U./Thurman, J.: „Themen und Herausforderungen der Regulierung von Suchmaschinen“. In: Machill, M./Beiler, M. (Hrsg.): *Die Macht der Suchmaschinen/The Power of Search Engines*. Köln: Herbert von Halem Verlag, 2007, S. 44-64.

Goldman 2006

Goldman, E.: „Search Engine Bias and the Demise of Search Engine Utopianism.“ In: *Yale Journal of Law and Technology* 9 (2006), S. 111-123.

Google 2003

Google: *Google News Deutschland (BETA)*. URL: http://news.google.de/intl/de_de/about_google_news.html [Stand: 2003].

Google 2006a

Google: *Google Annual Report 2005*. URL: http://investor.google.com/pdf/2005_Google_AnnualReport.pdf [Stand: 2006].

Google 2006b

Google: *Was ist Google AdSense?* URL: http://www.google.de/services/adsense_tour [Stand: 2006].

Green 2000

Green, D. C.: „*The evolution of Web searching*“. In: *Online Information Review* (2000), 24 (2): 124-137.

Griesbaum 2004

Griesbaum, J.: „Evaluation of three German search engines: Altavista.de, Google.de and Lycos.de.“ In: *Information Research* 9 (2004), Nr. 4. URL: <http://InformationR.net/ir/9-4/paper189.html> [Stand: 2007].

Gyongyi/Garcia-Molina 2005

Gyongyi, Z./Garcia-Molina, H.: *Web Spam Taxonomy*. URL: <http://airweb.cse.lehigh.edu/2005/gyongyi.pdf> [Stand: 2005].

Heise online 2003

Heise online: *Yahoo schluckt Overture*. URL: <http://www.heise.de/newsticker/meldung/38549> [Stand: 2003].

Heise online 2004

Heise online: *Bundestag sorgt sich um Vielfalt im Suchmaschinenmarkt*. URL: <http://www.heise.de/newsticker/meldung/48398> [Stand: 2004].

Heise online 2005

Heise online: *Frankreich forciert Pläne für europäische Google-Konkurrenz*. URL: <http://www.heise.de/newsticker/result.xhtml?url=/newsticker/meldung/63429&words=Quaero> [Stand: 2005].

Heise online 2006a

Heise online: *Google zensiert seine neue chinesische Suchmaschine*. URL: <http://www.heise.de/newsticker/meldung/68792> [Stand: 2006].

Heise online 2006b

Heise online: *Google sperrt nun auch deutsche Webseiten mit versteckten Suchwörtern aus*. URL: <http://www.heise.de/newsticker/meldung/69230> [Stand: 2006].

Heise online 2006c

Heise online: *US-Tageszeitungen lassen Werbeplatz in Printausgaben von Google vermarkten*. URL: <http://www.heise.de/newsticker/meldung/80560> [Stand: 2006].

Heise online 2006d

Heise online: *Gemeinsamer Sitemaps-Standard von Google, Microsoft und Yahoo*. URL: <http://www.heise.de/newsticker/meldung/81129> [Stand: 2006].

Heise online 2006e

Heise online: *400 Millionen Euro für europäische Suchmaschine Quaero*. URL: <http://www.heise.de/newsticker/meldung/77774> [Stand: 2006].

Heise online 2006f

Heise online: *Google: „Wir bemühen uns sehr um Datenschutz“*. URL: <http://www.heise.de/newsticker/meldung/79762> [Stand: 2006].

Heise online 2006g

Heise online: *Google verhandelt über Filmrechte für YouTube*. URL: <http://www.heise.de/newsticker/meldung/80463> [Stand: 2006].

Höhne/Ruß-Mohl 2005

Höhne, A./Ruß-Mohl, S.: „Schwarze Schafe in medialen Grauzonen. Ökonomische Analyse journalistischen Fehlverhaltens.“ In: *Neue Züricher Zeitung* 71 (2005), S. 50.

Hölscher 2002

Hölscher, C.: *Die Rolle des Wissens im Internet. Gezielt suchen und kompetent auswählen*. Stuttgart: Klett Verlag, 2002.

IAB 2006

IAB: *Internet advertising revenues grow 30% to a record \$ 12.5 billion in '05*. URL: http://www.iab.net/news/pr_2006_04_20.asp [Stand: 2006].

iBusiness 2006

iBusiness: *Ausgaben für Suchmaschinenmarketing verdoppeln sich in 2006*. URL: <http://www.ibusiness.de/aktuell/db/1138875019.html> [Stand: 2006].

Introna/Nissenbaum 2000

Introna, L./Nissenbaum, H.: „Shaping the Web: Why the Politics of Search Engines Matters.“ In: *The Information Society* 3 (2000), S. 1-17.

Jansen/Pooch 2001

Jansen, B./Pooch, U.: „Web user studies: A review and framework for future work.“ In: *Journal of the American Society of Information Science and Technology* 52 (2001), S. 235-246.

Jiang/Joshi/Krishnapuram/Yi 2000

Jiang, Z./Joshi, A./Krishnapuram, R./Yi, L.: *Retriever: Improving Web Search Engine Results Using Clustering*. URL: <http://www.cs.umbc.edu/~ajoshi/web-mine/retriever.ps.gz> [Stand: 2000].

Jones 2005

Jones, T.: „Both Sides of the Digital Battle for a High Rank from a Search Engine.“ In: *Association for Computing Machinery New Zealand Bulletin* 1 (2005), Nr. 2. URL: <http://www.cs.otago.ac.nz/postgrads/andrew/jones-2005-1.pdf> [Stand: 2007].

Keel/Bernet 2005

Keel, G./Bernet, M.: *Journalisten im Internet 2005. Eine Befragung von Deutschschweizer Medienschaffenden zum beruflichen Umgang mit dem Internet*. URL: http://www.iam.zhwin.ch/download/Studie_2005.pdf [Stand: 2005].

Keel/Wyss 2006

Keel, G.; V. Wyss: „Impact of journalists' Internet routines on journalism quality“. In: *Die wachsende Macht von Suchmaschinen im Internet: Auswirkungen auf User, Medienpolitik und Medienbusiness*. Friedrich-Ebert-Stiftung, Lehrstuhl für Journalistik II der Universi-

tät Leipzig und Lehrstuhl für Internationalen Journalismus der Universität Dortmund, Berlin, 26.-27. Juni 2006. URL: http://www.uni-leipzig.de/journalistik/suma/pdf/KEEL_WYSS_serach_berlin_en.pdf [Stand: 2006].

Keith 2007

Keith, S.: „Searching for News Headlines: Connections between Unresolved Hyperlinking Issues and a New Battle over Copyright Online.“ In: Machill, M./Beiler, M. (Hrsg.): *Die Macht der Suchmaschinen / The Power of Search Engines*. Köln: Herbert von Halem Verlag, 2007, S. 202-219.

Kink/Hess 2007

Kink, N./Hess, T.: „Suchmaschinen als Substitut traditioneller Medien? Erste Ergebnisse einer Studie zum Wandel der Informationsbeschaffung durch Suchmaschinen.“ In: Machill, M./Beiler, M. (Hrsg.): *Die Macht der Suchmaschinen / The Power of Search Engines*. Köln: Herbert von Halem Verlag, 2007, S. 298-307.

Köver 2006

Köver, C.: „Ruckzuck gescannt“. In: *Die Zeit* 46 (9.11.2006), S. 38.

Lawrence/Giles 1999

Lawrence, S./Giles, C. L.: „Accessibility of Information on the web.“ In: *Nature* 400 (1999), Nr. 8, S. 107-109.

Lewandowski 2007

Lewandowski, D.: „Mit welchen Kennzahlen lässt sich die Qualität von Suchmaschinen messen?“ In: Machill, M./Beiler, M. (Hrsg.): *Die Macht der Suchmaschinen / The Power of Search Engines*. Köln: Herbert von Halem Verlag, 2007, S. 243-258.

Lewandowski/Wahlig/Meyer-Bautor 2006

Lewandowski, D./Wahlig, H./Meyer-Bautor, G.: „The Freshness of Web Search Engine Databases.“ In: *Journal of Information Science*, 32 (2006), Nr. 2, S. 133-150.

Lünenbürger-Reidenbach/Petersen/Wagensonner 2000

Lünenbürger-Reidenbach, W./Petersen, J./Wagensonner, T.: *Media Studie 2000. Journalisten online*. URL: <http://www.newsaktuell.de/de/download/ms2000-d.pdf> [Stand: 2002].

Machill 2005

Machill, M.: *12 goldene Suchmaschinen-Regeln*. URL: http://www.lfm-nrw.de/downloads/LFM_Ratgeber.pdf [Stand: 2005].

Machill/Beiler 2005

Machill, M./Beiler, M.: „Qualität und Perspektiven der online-journalistischen Ausbildung an Hochschulen.“ In: Fasel, C. (Hrsg.): *Qualität und Erfolg im Journalismus*. Konstanz: UVK, 2005, S. 227-244.

Machill/Beiler 2006

Machill, M./Beiler, M.: „Internet-Suchmaschinen als neue Herausforderung für die Medienregulierung: Jugendschutz und publizistische Vielfalt als Fallbeispiele für Governance Issues.“ In: *Von der Medienpolitik zur Media Governance? Neue Problemstellungen, Ansätze und Formen der Regulierung öffentlicher Kommunikation*, Zürich, 16.-18.02.2006 (Tagung). URL: http://www.mediapolicy.unizh.ch/tagungen/mediagovernance/machill_beiler.pdf [Stand: 2006].

Machill/Beiler 2007

Machill, M./Beiler, M.: (Hrsg.): *Die Macht der Suchmaschinen / The Power of Search Engines*. Köln: Herbert von Halem Verlag, 2007.

Machill/Beiler/Neumann 2007

Machill, M./Beiler, M./Neumann, U.: „Leistungsfähigkeit von wissenschaftlichen Suchmaschinen. Ein Experiment am Beispiel von Google Scholar.“ In: Machill, M./Beiler, M. (Hrsg.): *Die Macht der Suchmaschinen / The Power of Search Engines*. Köln: Herbert von Halem Verlag, 2007, S. 327-347.

Machill/Lewandowski/Karzauninkat 2005

Machill, M./Lewandowski, D./Karzauninkat, S.: „Journalistische Aktualität im Internet. Ein Experiment mit den ‚News-Suchfunktionen‘ von Suchmaschinen.“ In: Machill, M./Schneider, N. (Hrsg.): *Suchmaschinen: Neue Herausforderungen für die Medienpolitik*. Berlin: Vistas Verlag, 2005, S. 105-164.

Machill/Neuberger/Schindler 2002

Machill, M./Neuberger, C./Schindler, F.: *Transparenz im Netz. Funktionen und Defizite von Internet-Suchmaschinen*. Gütersloh: Verlag Bertelsmann Stiftung, 2002.

Machill/Neuberger/Schweiger/Wirth 2003

Machill, M./Neuberger, C./Schweiger, W./Wirth, W.: „Wegweiser im Netz: Qualität und Nutzung von Suchmaschinen.“ In: Machill, M./Welp, C. (Hrsg.): *Wegweiser im Netz*. Gütersloh: Verlag Bertelsmann Stiftung, 2003, S. 13-490.

Marchionini 1995

Marchionini, G.: *Information seeking in electronic environments*. Cambridge, MA: Cambridge University Press, 1995.

Meier 1997

Meier, K.: *Experten im Netz. Maklersysteme als Recherchehilfe für Journalisten im Wissenschaftsbereich*. Konstanz: UVK-Medien, 1997.

Meier 2002

Meier, K.: „Grundlagen journalistischer Recherche im Internet.“ In: Meier, K. (Hrsg.): *Internet-Journalismus*. Konstanz: UVK, 2002, S. 297-357.

Meier 2004

Meier, K.: „Qualität im Online-Journalismus.“ In: Bucher, H.-J./Altmeyen, K.-D. (Hrsg.): *Qualität im Journalismus. Grundlagen – Dimensionen – Praxismodelle*. Wiesbaden: Westdeutscher Verlag 2004, S. 247-266.

Middleberg/Ross 2000

Middleberg, D./Ross, S.: *The Seventh Annual Print and Broadcast Survey of Media in the Wired World*. URL: http://www.univie.ac.at/stalzer/Online-Mafo/Medien/Seventh_Annual_Media_in_the_Wir.pdf [Stand: 2000].

Middleberg/Ross 2006

Middleberg, D./Ross, S.: *The Eleventh Annual Euro RSCG Magnet and Columbia University Survey of the Media. Rebuilding Trust: Rebuilding Credibility in the Newsroom and Boardroom*. URL: <http://magnet.mediaroom.com/file.php/binaries/31/RebuildingTrust.pdf> [Stand: 2006].

Miller/Galanter/Pribram 1973

Miller, G. A./Galanter, E./Pribram, K. H.: *Strategien des Handelns*. Stuttgart: Klett Verlag, 1973.

Netzeitung 2006

Netzeitung: *Microsofts Milliardenangriff auf Google*. URL: <http://www.netzeitung.de/internet/396504.html> [Stand: 2006].

Netzwerk Recherche 2006

Netzwerk Recherche: „Mehr Qualität durch Recherche – Von der Pflicht zur Kür.“ In: *5. Jahreskonferenz netzwerk recherche*, Hamburg (Tagung). URL: <http://www.recherche-info.de> [Stand: 2006].

News aktuell 2000

News aktuell: *media Studie 2000. Journalisten online*. URL: <http://www.newsaktuell.de/mediaevents/mediastudie.htx> [Stand: 2000].

News aktuell 2002

News aktuell: *Media Studie 2002. Journalisten online – die Folgestudie*. URL: <http://www.newsaktuell.de/de/mediaevents/mediastudie.htx> [Stand: 2002].

Nielsen Netratings 2006

Nielsen Netratings: *Online searches grow 55 percent year-over-year to nearly 5.1 billion searches in Dezember 2005*. URL: http://www.nielsen-netratings.com/pr/pr_060209.pdf [Stand: 2006].

Orlowski 2003

Orlowski, A.: *Google News: press releases are OK – Official*. URL: http://www.theregister.co.uk/2003/04/05/google_news_press_releases [Stand: 2003].

Plog 2006

Plog, J.: *Rede anlässlich des Abends der Begegnung der Evangelisch-Lutherischen Landeskirche Braunschweig*. URL: http://www.landeskirche-braunschweig.de/luth-bs/wcms/uploads/download/AdB_Plog.pdf [Stand: 2006].

Rieder 2005

Rieder, B.: „Networked Control: Search Engines and the Symmetry of Confidence.“ In: *International Review of Information Ethics* 3 (2005), S. 3-17.

Rötzer 2005

Rötzer, F.: „Vom Begriff ‚Versager‘ zur Biographie von G. W. Bush.“ In: *Telepolis*. URL: <http://www.heise.de/tp/r4/artikel/20/20240/1.html> [Stand: 2005].

Rötzer 2006

Rötzer, F.: „Wie man trotz aktiver Beihilfe zur Zensur für die Freiheit kämpft.“ In: *Telepolis*. URL: <http://www.heise.de/tp/r4/artikel/22/22042/1.html> [Stand: 2006].

Rotenberg 2007

Rotenberg, B.: „Towards Personalised Search: EU Data Protection Law and Implications for Media Pluralism.“ In: Machill, M./Beiler, M. (Hrsg.): *Die Macht der Suchmaschinen / The Power of Search Engines*. Köln: Herbert von Halem Verlag, 2007, S. 87-104.

Sadrozinski 2007

Sadrozinski, J.: „Suchmaschinen und öffentlich-rechtlicher Onlinejournalismus am Beispiel tagesschau.de.“ In: Machill, M./Beiler, M. (Hrsg.): *Die Macht der Suchmaschinen / The Power of Search Engines*. Köln: Herbert von Halem Verlag, 2007, S. 171-184.

Schindler 2005

Schindler, F.: „Suchmaschinen und Jugendschutz.“ In: Machill, M./Schneider, N. (Hrsg.): *Suchmaschinen: Neue Herausforderungen für die Medienpolitik*. Berlin: Vistas Verlag, 2005, S. 55-71.

Shneiderman/Byrd/Croft 1997

Shneiderman, B./Byrd, D./Croft, W. B.: „Clarifying search. A user-interface framework for text searches.“ In: *D-Lib Magazine* 3 (1997), Nr. 1. URL: <http://www.dlib.org/dlib/january97/retrieval/01shneiderman.html> [Stand: 2007].

Schmitz 2006

Schmitz, H.: „Kaum genutzte Transparenz. Journalisten und Informationsfreiheitsgesetze.“ In: *epd medien* 12 (2006), S. 3-7.

Schotzger 2005

Schotzger, E.: *Google ist wertvollstes Medienunternehmen der Welt*. URL: <http://www.presetext.com/pte.mc?pte=050609036> [Stand: 2005].

Schulz/Held 2007

Schulz, W./Held, T.: „Der Index auf dem Index? Selbstzensur und Zensur von Suchmaschinen.“ In: Machill, M./Beiler, M. (Hrsg.): *Die Macht der Suchmaschinen / The Power of Search Engines*. Köln: Herbert von Halem Verlag, 2007, S. 71-86.

Schulz/Held/Laudien 2005

Schulz, W./Held, T./Laudien, A.: *Suchmaschinen als Gatekeeper in der öffentlichen Kommunikation. Rechtliche Anforderungen an Zugangsoffenheit und Transparenz bei Suchmaschinen im WWW*. Berlin: Vistas Verlag, 2005.

Sonnenreich 1997

Sonnenreich, W.: *A History of Search Engines*. URL: <http://www.wiley.com/legacy/computers/books/sonnenreich/history> [Stand: 1997].

Spiegel Online 2006a

Spiegel Online: *Zensur im Internet. „Yahoo ist am schlimmsten“*. URL: <http://www.spiegel.de/netzwelt/politik/0,1518,413805,00.html> [Stand: 2006].

Spiegel Online 2006b

Spiegel Online: *Google Marktwert steigt über 150 Milliarden Dollar*. URL: <http://www.spiegel.de/wirtschaft/0,1518,444288,00.html> [Stand: 2006].

Spiegel Online 2006c

Spiegel Online: *Googles risikoreichste Investition*. URL: <http://www.spiegel.de/wirtschaft/0,1518,441686,00.html> [Stand: 2006].

Stöcker/Darnbeck 2006

Stöcker, C./Darnbeck, H.: *Quaero ist geplatzt*. URL: <http://www.spiegel.de/netzwelt/web/0,1518,455558,00.html> [Stand: 2006].

Sullivan 2006a

Sullivan, D.: *Search Engine Ratings and Reviews*. URL: <http://searchenginewatch.com/showPage.html?page=reports> [Stand: 2006].

Sullivan 2006b

Sullivan, D.: *Which Search Engines Log IP Addresses & Cookies – And Why Care?* Online: <http://blog.searchenginewatch.com/blog/060206-150030> [Stand: 2006].

TimeWarner 2006

TimeWarner: *Form 10K-A. Amended Annual Report*. URL: <http://www.shareholder.com/Common/Edgar/1105705/950144-06-8743/06-00.pdf> [Stand: 2006].

Trappel/Meier/Schrabe/Wölk 2002

Trappel, J./Meier, W. A./Schrabe, K./Wölk, M.: *Die gesellschaftlichen Folgen der Medienkonzentration*. URL: <http://www.lfm-nrw.de/downloads/studie-medienkonzentration.pdf> [Stand: 2002].

Van Couvering 2004

Van Couvering, E.: *New Media? The Political Economy of Internet Search Engines*. URL: http://personal.lse.ac.uk/vancouve/IAMCR-TP_SearchEnginePoliticalEconomy_EVC_2004-07-14.pdf [Stand: 2004].

Van Couvering 2007

Van Couvering, E.: „The Economy of Navigation: Search Engines, Search Optimisation and Search Results.“ In: Machill, M./Beiler, M. (Hrsg.): *Die Macht der Suchmaschinen / The Power of Search Engines*. Köln: Herbert von Halem Verlag, 2007, S. 105-126.

Van Eimeren/Frees 2006

Van Eimeren, B./Frees, B.: „ARD/ZDF-Online-Studie: Schnelle Zugänge, neue Anwendungen, neue Nutzer?“ In: *Media Perspektiven* 8 (2006), S. 402-415.

Vaughan/Thelwall 2004

Vaughan, L./Thelwall, M.: „Search Engine Coverage Bias: Evidence and Possible Causes.“ In: *Information Processing & Management* 40 (2004), Nr. 4, S. 693-707.

Véronis 2006

Véronis, J.: *A comparative study of six search engines*. URL: <http://www.up.univ-mrs.fr/veronis/pdf/2006-comparative-study.pdf> [Stand: 2006].

Vine 2004

Vine, R.: *The Business of Search Engines. Understanding how Web advertising, partnerships, and the race for market dominance affect search tools and search results*. URL: http://www.workingfaster.com/2004_business_of_search_engines_final.pdf [Stand: 2004].

Weber/Groner 1999

Weber, C./Groner, R.: „Suchstrategien im WWW bei Laien und Experten.“ In: Wirth, W./Schweiger, W. (Hrsg.): *Selektion im Internet. Empirische Analysen zu einem Schlüsselkonzept*. Opladen: Westdeutscher Verlag, 1999, S. 181-198.

Wegner 2002

Wegner, J.: „Der neue Google-Faktor.“ In: *message werkstatt* 1 (2002), S. 2-3.

Wegner 2005

Wegner, J.: „Die Welt laut Google. Oder: Was heisst hier Recherche? Eine viel zu schnelle Suche.“ In: *epd medien* (19.1.2005), S. 8-11.

Weischenberg/Scholl/Malik 2006

Weischenberg, S./Scholl, A./Malik, M.: „Journalismus in Deutschland 2005. Zentrale Befunde der aktuellen Repräsentativbefragung deutscher Journalisten.“ In: *Media Perspektiven* 7 (2006), S. 346-361.

Welp 2003

Welp, C.: „Ein Code of Conduct für Suchmaschinen.“ In: Machill, M./Welp, C. (Hrsg.): *Wegweiser im Netz*. Gütersloh: Verlag Bertelsmann Stiftung, 2003, S. 491-508.

Winkler 1997

Winkler, H.: „Suchmaschinen. Metamedien im Internet.“ In: *Telepolis*. URL: <http://www.telepolis.de/r4/artikel/1/1135/1.html> [Stand: 1997].

Winterbauer 2003

Winterbauer, S.: „Das Googlepol.“ In: *Süddeutsche Zeitung* (19.7.2003), S. 18.

Wirth/Schweiger 1999

Wirth, W./Schweiger, W.: „Selektion neu betrachtet: Auswahlentscheidungen im Internet.“ In: Wirth, W./Schweiger, W. (Hrsg.): *Selektion im Internet. Empirische Analysen zu einem Schlüsselkonzept*. Opladen: Westdeutscher Verlag, 1999, S. 43-74.

Wolling 2002

Wolling, J.: „Suchmaschinen – Gatekeeper im Internet.“ In: *Medienwissenschaft Schweiz* 2 (2002), S. 15-23.

Yahoo 2006

Yahoo: *Yahoo! Annual Report 2005*. URL: <http://www.shareholder.com/shared/dynamicdoc/YH00/1183/YH00.pdf> [Stand: 2006].

> VON SUCHMASCHINEN ZU ANTWORTMASCHINEN: SEMANTISCHE TECHNOLOGIEN UND BENUTZERPARTI- ZIPATION IM WEB 3.0

WOLFGANG WAHLSTER

1 EINLEITUNG

Die Suchmaschinen von morgen werden klar über die Leistungen der heute führenden Technologie von Google hinausgehen, indem sie nicht nur die Eingabe von Schlüsselwörtern, sondern komplexe natürlichsprachlich formulierte Anfragen zulassen und Suchergebnisse höherer Qualität liefern, die zudem an den aktuellen Kontext und die jeweilige Aufgabenstellung des Benutzers angepasst sind. Sie werden Anfragen des Benutzers inhaltlich verstehen und statt einer Vielzahl von Verweisen auf Dokumente, in denen möglicherweise die gesuchte Information enthalten ist, gleich die Antwort auf die Benutzeranfrage liefern. So werden aus Suchmaschinen in Zukunft Antwortmaschinen, weil nicht mehr die Unterstützung eines Suchprozesses, sondern die rasche Generierung einer gezielten Antwort auf eine konkrete Anfrage die Systemleistung ausmacht (vgl. Abb. 1).

Abb. 1: Stichwortbasierte Suche versus Fragebeantwortung

Heute: Google

• **Eingabe:** Umsatz von Siemens im Jahr 2005 in Deutschland



• **Ausgabe:**

[Fritz/Fischer - Investor - FAT MET - Medizinische Porsche...](#)
Der Umsatz legte um sechs Prozent auf 12,64 Milliarden Dollar zu ... die um Siemens-Cassen rund 200 Millionen Euro abgezweigt und die Summe als ...
[www.faz.net / Doc--E60DEA220797A43A3996599C4937E112A-A7Jh-Ezomwv-Scaetent.html - 62k ...](#)
[Siemens Financial Services: „Leasingfabrik“ als neue...](#)
Der Umsatz legte um sechs Prozent auf 12,64 Milliarden Dollar zu ... für die USA) hatten dafür zwischen Februar und April 2005 in Deutschland, Frankreich, ...
[www.silicon.de/arsip_services/14407 - 32k - In Cache - Ähnliche Seiten](#)

Morgen: SmartWeb

• **Eingabe:**  Umsatz von Siemens im Jahr 2005 in Deutschland?

• **Ausgabe:**  15,685 Milliarden Euro

Siemens	2004	2005	2006
Siemens (ohne den Bereich ITC)	24.612	23.767	23.094
ITC	14.000	14.076	14.000
Siemens (ohne ITC)	8.224	7.542	7.000
Siemens ITC	10.007	8.400	8.012
Siemens (ohne ITC)	5.174	4.500	4.000
Siemens	70.488	70.207	70.207

2 VON SYNTAKTISCHEN SUCHMASCHINEN ZU SEMANTISCHEN ANTWORTMASCHINEN

Besonders in mobilen Anwendungssituationen will der Benutzer als Ausgabe keine Listen von Verweisen auf möglicherweise relevante Webdokumente, die er dann wiederum weiter durchsuchen muss, sondern eine konzise Antwort auf seine Anfrage. So will der Familienvater während einer Autofahrt, wenn das Kind auf dem Rücksitz plötzlich über starke Schmerzen klagt, auf seine Frage: „Wo ist der nächste Kinderarzt?“, nur die Antwort: „Dr. Pfeifer in der Marienstraße 23“, sodass er als Folgeauftrag das System bitten kann: „Dann zeige mir die Route und führe mich dort hin“, um eine Navigationsunterstützung zum nächsten Kinderarzt zu erhalten. In dem vom Bundesministerium für Bildung und Forschung (BMBF) von 2004 bis 2007 mit 13,7 Mio. Euro geförderten Verbundprojekt SmartWeb wurden unter meiner Leitung die Grundlagen für semantische Antwortmaschinen erarbeitet und erstmals eine solche Funktionalität für den mobilen Benutzer vollständig realisiert. In einem Mercedes-Testfahrzeug konnten sich die Besucher der CeBIT 2007 davon überzeugen, dass solche mobilen Antwortmaschinen heute schon in der Lage sind, erfolgreich Dialoge wie die als Beispiel angeführte Suche nach einem Kinderarzt mit dem Fahrer durchzuführen.

SmartWeb¹ zeigt damit den Weg auf, wie folgende Defizite bisheriger Suchmaschinen überwunden werden können:

- Die Anfrage des Benutzers wird nicht inhaltlich verstanden, sondern nur als Zeichenkette betrachtet.
- Jede Anfrage wird isoliert beantwortet, ohne den Inhalt vorangegangener Anfragen oder Suchergebnisse und damit den Dialogkontext zu berücksichtigen.
- Suchanfragen, die sich auf mehrstellige Relationen zwischen Begriffen beziehen und komplexe Restriktionen enthalten, führen zu unbefriedigenden oder sogar falschen Resultaten.
- „Versteckte Inhalte“, die im sogenannten „Deep Web“ nur über Portale, Webdienste oder in PDF-Dokumenten zugänglich sind, werden nicht gefunden.

Nutzer traditioneller Web-Suchmaschinen klagen wegen dieser Defizite oftmals über irrelevante Ergebnisse und einen zu umfangreichen Ergebnisraum. Sie verlassen die Suchmaschine, ohne eine Antwort auf ihre Ursprungsfrage erhalten zu haben, und versuchen weitere Recherchen mit anderen Suchmaschinen. SmartWeb (vgl. Abb. 2) bietet erstmals in einem integrierten Gesamtsystem alle Voraussetzungen, um diese Defizite zu überwinden und so von traditionellen Suchmaschinen zu mobilen Antwortmaschinen zu kommen.

¹ Wahlster 2008.

Abb. 2: SmartWeb als mobile Antwortmaschine über die FIFA WM 2006



Ein anschauliches Beispiel für die Leistungsfähigkeit von SmartWeb als Antwortmaschine entstand bei einem Blindtest durch den für die Förderung von SmartWeb verantwortlichen Leiter des Referats 524 des BMBF, Rainer Jansen. Obwohl SmartWeb vielfach in der Suchdomäne „Sport“ und speziell „FIFA Fußballweltmeisterschaft 2006“ getestet war (vgl. Abb. 2), stellte Herr Jansen eine Frage aus dem Kulturbereich (siehe Abb. 3), die komplexe Relationen (Name von Sänger/in in der Titelrolle von La Traviata) und zwei Restriktionen (bei den Salzburger Festspielen, in der Premiere) enthielt. Obwohl eine ähnliche Frage SmartWeb zuvor noch nie gestellt worden war, wurde zur Verblüffung auch der anwesenden Fachwissenschaftler diese schwierige Frage präzise und korrekt mit „Anna Netrebko“ beantwortet.

Abb. 3 zeigt auch ein Beispiel dafür, dass der SmartWeb-Nutzer eine bestimmte Modalität der Antwort anfordern kann, indem er am unteren Rand der Bildschirmmaske von SmartWeb auf eines der vier Ikonen für Bild, Videoclip, Text- oder Tondokument mit dem Stift zeigt. Im vorliegenden Fall wird dann zusätzlich zu der Standardantwort in textueller Form noch ein Bild der Künstlerin ausgegeben, das durch eine Bildsuche im Web gefunden wurde.

Abb. 3: Beantwortung einer komplexen Suchanfrage im offenen Web durch SmartWeb



Über alle Systemversionen hinweg bietet SmartWeb seinem Benutzer einen symmetrischen multimodalen Zugriff,² der auf der Eingabeseite neben gesprochener und geschriebener Sprache auch Gestik, Haptik und Video kombiniert mit Sprache, Haptik, Grafik, Video und Ton auf der Ausgabeseite. Der Benutzer bekommt die Informationen also nicht nur durch Eintippen von Fragen auf der kleinen Tastatur des mobilen Endgeräts, sondern auch durch gesprochene Sprache, zum Beispiel kombiniert mit Stifteingaben auf dem Bildschirm. Es wird also möglich, das Mobiltelefon in Alltagssprache zu „fragen“, und dieses wird mit Informationen aus dem Netz „antworten“. SmartWeb baut auf das BMBF-Vorgängerprojekt SmartKom (1999 bis 2003) auf, das erstmals das Prinzip der symmetrischen Multimodalität beim semantischen Informationszugriff realisierte,³ allerdings im Gegensatz zu SmartWeb keinen freien Zugang zum offenen Web erlaubte, sondern ausschließlich in vorher festgelegten Diskursbereichen funktionierte.

SmartWeb hat mit 16 Partnern aus Industrie und Forschung wesentliche wissenschaftliche und technische Grundlagen für das neue Leuchtturmprojekt *Theseus* gelegt, das im Rahmen der Hightech-Strategie der Bundesregierung vom Bundesministerium für Wirtschaft und Technologie (BMWi) zunächst mit 90 Mio. Euro ab 2007 für fünf Jahre gefördert wird. In *Theseus* sollen von 30 Konsortialpartnern die semantischen Technologien für innovative Anwendungen des zukünftigen Internets der Dienste erschlossen werden.

Das Internet der Dienste wird die automatische Komposition von Webdiensten ermöglichen, die zur Beantwortung komplexer Benutzeranfragen notwendig wird. Fragt beispielsweise eine Sekretärin während der Vorbereitung der nächsten Geschäftsreise ihres Chefs nach Japan: „Wieviel Euro kostet die schnellste und preiswerteste Verbindung vom Flughafen Narita zum Palace Hotel in Tokio?“, dann sollte sie zukünftig auf der Basis der *Theseus*-Technologie die Antwort erhalten: „Mit dem Narita-Express kostet dies derzeit 16,60 Euro.“ Dazu muss die semantische Antwortmaschine nicht nur die Preise verschiedener Bahn- und Busverbindungen und deren Fahrzeiten auf deren jeweiligen Webportalen finden, extrahieren und vergleichen, sondern auch die Kosten von 2.700 Yen in Euro umrechnen. Dazu muss ein passender Webdienst wie der Universal Currency Converter automatisch gefunden und mit den entsprechenden Eingabeparametern versorgt werden, sodass nach dem aktuellen Umrechnungskurs die 16,60 Euro als Preis ausgegeben werden können. Wichtig ist auch die Dialogfähigkeit solcher Systeme, die sicherstellt, dass eine Folgefrage wie: „Wann gibt es den passenden Anschluss zum Lufthansa-Flug nach Tokio?“ im Kontext der Reiseplanung richtig verstanden wird, sodass „16.40, Gleis 4“ für den Narita-Express bestimmt wird. Da semantische Antwortmaschinen das gesamte Web als Wissensbasis nutzen, ist natürlich kritisch zu hinterfragen, ob man den Antworten vertrauen kann oder ob hier Fehlinformation aus unzuverlässigen

² Wahlster 2003.

³ Wahlster 2006.

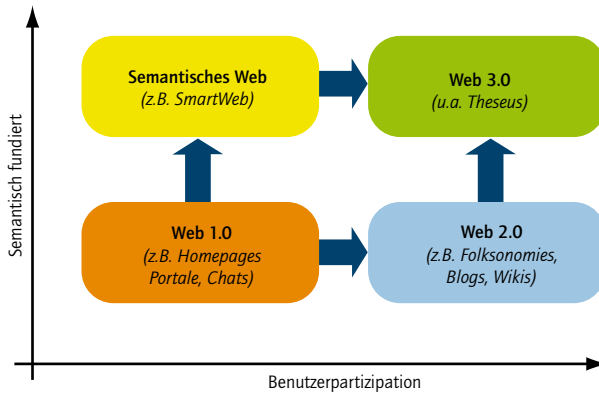
Webseiten verarbeitet wurde. Daher müssen zukünftige Antwortmaschinen auch eine Erklärungskomponente enthalten, die es dem Benutzer erlaubt, die Informationsquellen und Webdienste kritisch zu prüfen, auf denen die Systemantwort beruht. In unserem Beispiel würde die Antwortmaschine den kritischen Nutzer darauf verweisen, dass die Information von der offiziellen und aktuellen Homepage der Japanischen Eisenbahnen stammt und der verwendete Währungsumrechner im Internet von einer seit 1993 zuverlässig arbeitenden Internetfirma XE.com betrieben wird.

3 WEB 3.0 = SEMANTISCHES WEB + WEB 2.0

Das syntaktische Web 1.0 hat multimediale Dokumente und Dienste im Internet miteinander vernetzt, wobei die Relationen zwischen diesen Objekten keine maschinenverstehbare Semantik hatten, sondern lediglich zur hypermedialen Navigation des Benutzers dienten. Das Web 1.0 hat zwar den weltweiten Zugang zu digital gespeicherter Information drastisch verbessert; seine Inhalte sind aber nur maschinenlesbar, ohne maschinell verstehbar zu sein. Da Information im Web meist in natürlicher Sprache präsentiert wird, sind die gefundenen Dokumente bislang nur für den Menschen voll verständlich. Das semantische Web⁴ basiert dagegen auf der inhaltlichen Beschreibung digitaler Dokumente mit standardisierten Vokabularien, die eine maschinell verstehbare Semantik haben. Damit wird der Übergang von einem „Netz aus Verweisstrukturen“ zu einem „Netz aus Inhaltsstrukturen“ vollzogen. Den Kern der semantischen Technologie bilden Markierungssprachen, die eine formale Syntax und Semantik haben und in Form einer Ontologie eine standardisierte Begrifflichkeit zur Beschreibung digitaler Inhalte bereitstellen. Semantische Markierungssprachen wie OWL erlauben die weltweite Verbreitung und gemeinsame Nutzung von Ontologien. Die semantische Markierung durch Metadaten stellt neben der Markierung für das Layout (HTML) und die Struktur (XML) eines Dokumentes die dritte Ebene einer Dokumentbeschreibung dar. Bei der Suche nach Informationen können Schlussfolgerungen den Suchprozess beschleunigen oder gesuchte Information ableiten. Methoden des maschinellen Lernens, des Data Mining und des Text Mining reduzieren beim Aufbau und bei der Wartung von Ontologien den manuellen Aufwand. Software-Werkzeuge unterstützen das Erstellen, Editieren, Visualisieren, Auswerten, Verschmelzen und Validieren von Ontologien. Durch Zugriff auf semantische Webseiten können intelligente Suchagenten mit hoher Präzision gezielt Informationen im Internet auffinden, filtern und kombinieren, um dem Benutzer zeitaufwendige Recherchearbeit abzunehmen.

⁴ Fensel/Hendler/Lieberman/Wahlster 2003 und Wahlster 2004.

Abb. 4: Das Web 3.0 als Kombination von Semantischem Web und Web 2.0



Noch sind semantische Suchmaschinen nicht universell einsatzfähig, weil die Zahl der bereits semantisch annotierten Webdokumente und Webdienste noch im einstelligen Prozentbereich liegt. Dies ist auf den hohen manuellen Aufwand bei der nachträglichen semantischen Annotation bestehender Webseiten zurückzuführen. Sprachtechnologische Verfahren zur automatischen semantischen Annotation von Webseiten werden zwar immer leistungsfähiger, erreichen aber nur selten – und dann in sehr speziellen Wissensbereichen – die Qualität und Zuverlässigkeit menschlicher Annotationen.

Es liegt daher nahe, das Prinzip der Benutzerpartizipation des Web 2.0 als Massenbewegung für den Aufbau des semantischen Webs zu nutzen. Im Web 2.0 ist der Konsument gleichzeitig auch Produzent (ein „Prosumer“) und daher motiviert, die von ihm mitgestalteten Wikis, Foto-, Video- und Musiksammlungen stets zu aktualisieren und zu pflegen. Wenn es gelingt, auch den nichtprofessionellen Benutzer aktiv dabei mitwirken zu lassen, semantisch fundierte Webseiten und Webdienste zu schaffen, entsteht das Web 3.0 (vgl. Abb. 4). Dabei werden ontologische Klassifikationen in einfache Interaktionssequenzen auf natürliche Weise eingebunden, sodass es auch für den informatisch nicht geschulten Nutzer möglich wird, sein Weltwissen nutzbringend beim Aufbau des Web 3.0 einzubringen, um somit die Voraussetzung für hochpräzise Antwortmaschinen zu schaffen. Diesen Ansatz der Kombination aus semantischem Web und Web 2.0 zum neuartigen Web 3.0 der nächsten Generation verfolgt das Theseus-Konsortium.

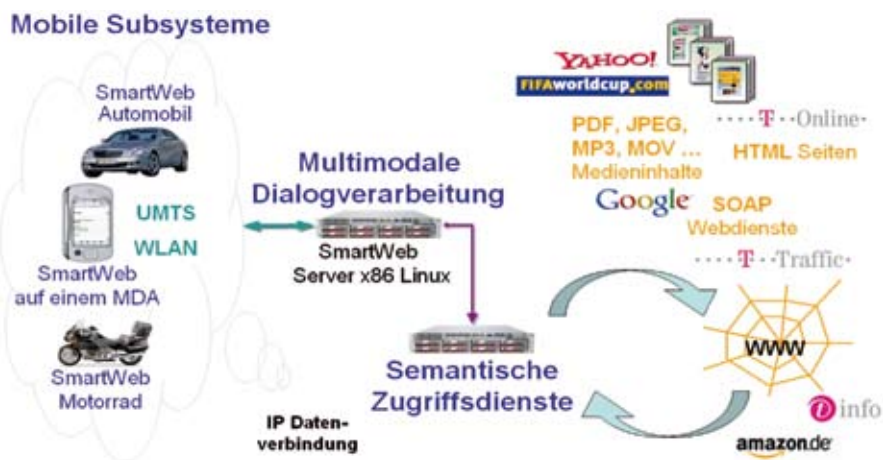
4 SMARTWEB: EINE MULTIMODALE UND MOBILE ANTWORTMASCHINE

SmartWeb besteht aus drei großen Softwarebereichen:⁵ den mobilen Subsystemen (für Mobiltelefone der MDA-Serie von T-Mobile, Autos und Motorräder), der multimodalen Dialogverarbeitung und den semantischen Zugriffsdiensten (vgl. Abb. 5). SmartWeb kann parallel eine Vielzahl von Benutzern unterstützen, weil das System nach dem

⁵ Reithinger/Herzog/Blocher 2007, Sonntag/Engel/Herzog/Pfalzgraf/Pfleger/Romanelli/Reithinger 2007 und Wahlster 2007.

Client-Server-Prinzip arbeitet und der Server gleichzeitig mehrere Dialoge mit verschiedenen mobilen Nutzern abwickeln kann. Dazu wird von der SmartWeb-Dialogplattform für jeden neuen eingehenden Anruf eine neue Instanz des Dialogservers instantiiert. Die Hauptverarbeitungsleistung von SmartWeb wird nicht auf dem mobilen Endgerät, sondern auf dem über eine breitbandige Funkverbindung erreichbaren Server erbracht, auf dem eine leistungsfähige Spracherkennung mit einer semantischen Anfrageanalyse und einer Komponente zum Dialogmanagement kombiniert ist. Auf dem MDA-Client läuft eine Java-basierte Steuerungskomponente für alle multimodalen Ein- und Ausgaben sowie die grafische Bedienungsoberfläche von SmartWeb. Falls einmal die Verbindung zum Sprachdialog-Server abbricht, verfügt der MDA zur Überbrückung der Störung auch über eine einfache Sprachbedienung auf der Basis von VoiceXML.

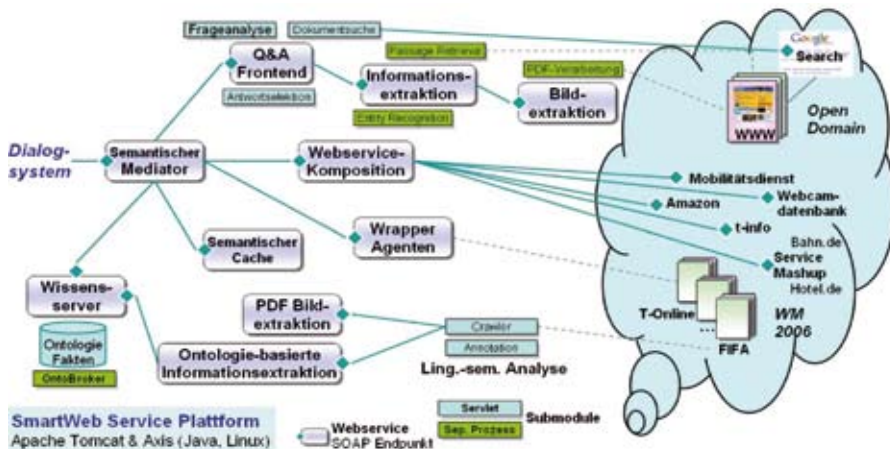
Abb. 5: Die drei großen Softwarebereiche von SmartWeb



Die Beantwortung von Anfragen an SmartWeb wird nicht nur über ein einziges Verfahren realisiert, sondern ein semantischer Mediator (vgl. Abb. 6) entscheidet, wie eine vom Dialogsystem analysierte Anfrage mithilfe der verschiedenen semantischen Zugriffsdienste von SmartWeb beantwortet werden soll. Neben der gezielten Informationsextraktion aus Passagen von Dokumenten im offenen Web, die zunächst mit klassischen Suchmaschinen wie Google gefunden werden, kann zum Beispiel auch eine Komposition von Webdiensten oder ein Aufruf eines Zugriffsagenten auf ein Portal durch den semantischen Mediator ausgelöst werden. Wie in Verbmobil⁶ wird in SmartWeb also ein Ansatz verfolgt, der mehrere Lösungswege parallel vorhält und je nach Aufgabenstellung den besten Ansatz zur Reaktion auf die Benutzereingabe verfolgt.

⁶ Wahlster 2000.

Abb. 6: Die Ansteuerung von Zugriffsdiensten durch den semantischen Mediator

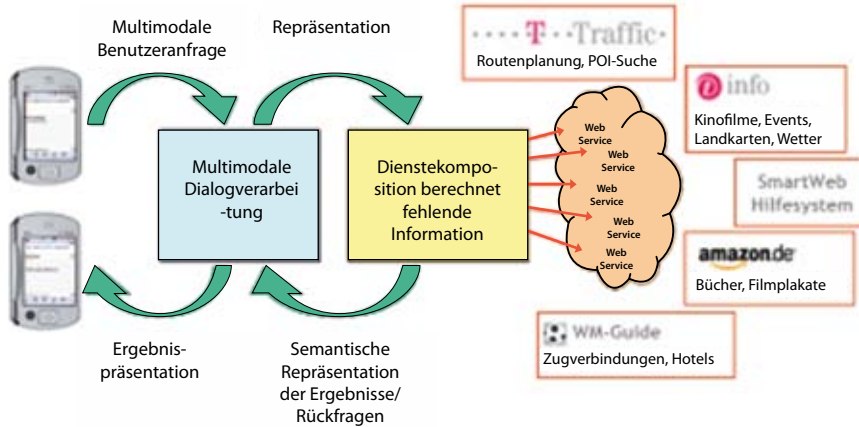


SmartWeb kann neun verschiedene Webdienste automatisch je nach Bedarf zu höherwertigen Diensten komponieren (unter anderem Routenplanung, Wetterinformation, Veranstaltungskalender, Verkehrsinformation, Webcam-Suche). Dabei nutzt das System die semantische Beschreibung dieser Webdienste mithilfe von Erweiterungen der standardisierten Dienstbeschreibungssprache OWL-S, die auf einer Beschreibungslogik mit effizienten Inferenzverfahren beruht. Die Komposition der Dienste erfolgt in SmartWeb über einen musterbasierten Planungsalgorithmus, der bei Unterspezifikation von Anfragen auch einen Klärungsdiallog mit dem Benutzer initiieren kann (vgl. Abb. 7). Ein wichtiges Entwurfsprinzip von SmartWeb war die gezielte Benutzung von Ontologien in allen Verarbeitungskomponenten.⁷ Die ontologische Infrastruktur SWIntO (SmartWeb Integrated Ontology) basiert auf einer Kombination der weitverbreiteten Universalontologien DOLCE und SUMO, die durch spezielle Diskurs- und Medienontologien ergänzt wurden.⁸ Der W3C-Standard EMMA (Extensible MultiModal Annotation Markup Language) wurde in SmartWeb so ausgebaut, dass nicht nur die Semantik der multimodalen Benutzereingaben, sondern auch die der multimodalen Systemausgaben formal damit beschrieben werden kann, sodass alle Komponenten auf dem Server Information im EMMA-Format austauschen können.

⁷ Oberle et al. 2007.

⁸ Ebenda.

Abb. 7: Planbasierte Dienstekomposition in SmartWeb



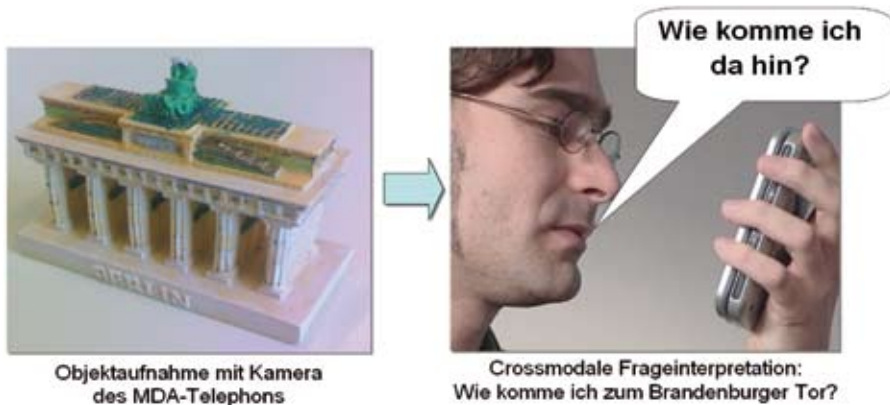
Da wir heute noch ganz am Anfang des semantischen Webs als zukünftiges Internet stehen und nur ganz wenige Webinhalte schon ontologisch annotiert sind, wurde in SmartWeb auch intensiv an der automatischen Generierung von semantischen Annotationen für bestehende HTML-Seiten, Fotos und PDF-Dokumente aus dem Web gearbeitet. Anhand eines Korpus zum Bereich „Fußball“ wurden 5.000 HTML-Seiten, 300 PDF-Dokumente und 15.000 Fotos mit sprachtechnologischen Methoden der Informationsextraktion⁹ sowie wissensbasierten Layout- und Bildanalyseverfahren bearbeitet. Für die Anwendungsdomäne „FIFA Fußballweltmeisterschaft 2006“ konnten im finalen Prototyp von SmartWeb 5.000 Fotos anhand der Bildunterschriften mit Spielen und Personen und 3.400 Textpassagen zu Spielereignissen automatisch semantisch annotiert werden. Mit 110 komplett annotierten Mannschaftsfotos, 40 markierten Videosequenzen und 12 Audioreportagen ergeben sich rund 100.000 Ontologieinstanzen in einer umfangreichen Wissensbasis, die manuell erstellte mit automatisch extrahierten semantischen Annotationen aller Medienobjekte kombiniert.

⁹ Cramer/Leidner/Klakow 2006.

5 MULTIMODALE UND LOKATIONSENSITIVE ANTWORTMASCHINEN

In SmartWeb kann die Handykamera der MDA-Telefone genutzt werden, um das Eintippen oder Sprechen komplexer Objektnamen bei Anfragen an das System zu umgehen. Der Benutzer fotografiert einfach das Referenzobjekt selbst, ein Foto oder ein Modell von ihm, und ein servergestütztes Objekterkennungssystem ordnet dann als Webdienst aufgrund charakteristischer Bildpunkte selbst bei geändertem Aufnahmewinkel und modifizierten Beleuchtungsparametern das Foto einem in einer Bilddatenbank abgelegten Referenzbild zu. Hier wird also erstmals eine Bildeingabe mit einer Spracheingabe kombiniert, sodass crossmodale Fragen wie in Abb. 8 möglich werden: Nachdem der Nutzer ein Modell des Brandenburger Tores mit seiner Handykamera fotografiert hat, kann er einfach fragen: „Wie komme ich da hin?“

Abb. 8: Crossmodale Anfragen durch Kombination von Handykamera und Spracheingabe



Eine weitere Innovation von SmartWeb ist die Möglichkeit zur Beantwortung von crossmodalen Anfragen, die sich auf Bildmaterial beziehen, welches als Ergebnis einer vorangegangenen Anfrage von SmartWeb auf dem MDA präsentiert wurde. SmartWeb nutzt bei einer Folgefrage wie: „Wer ist der dritte links oben?“ (vgl. Abb. 9) dazu die MPEG7-Repräsentation der Bildsegmente und interpretiert räumliche Relationen wie den Ausgangspunkt (links oben), um den referenzierten Spieler (den dritten) zu bestimmen.

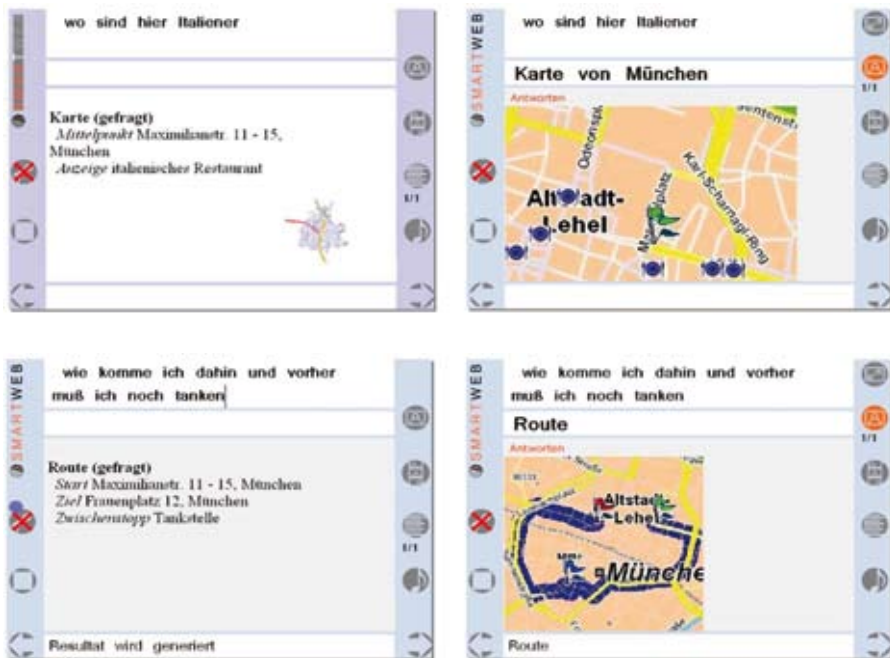
Abb. 9: Beantwortung von Anfragen mit crossmodaler räumlicher Referenz auf Bildmaterial



Der Spieler mit der Nummer 18 wird durch die semantische Annotation des Bildes als „Tim Borowski“ bestimmt, sodass die Antwort von SmartWeb ausgegeben werden kann. Die Bilder sind dazu dekomponiert und die einzelnen Bildsegmente werden mit Instanzen aus der Medienontologie annotiert, die wiederum mit einer Sportontologie verknüpft ist, in der zum Beispiel „Lehmann“ als Torwart klassifiziert ist.

Eine Weltneuheit bietet SmartWeb auch mit Sprachdialogen, in denen das System mehrere Internet-Dienste mit Fahrzeugdiensten automatisch verknüpft: Nachdem der Fahrer auf die lokationssensitive Frage: „Wo gibt es hier Italiener?“ entsprechende Restaurants auf der digitalen Karte angezeigt bekommen und eines ausgesucht hat, kann er den frei formulierten Dialog fortsetzen mit: „Wie komme ich dahin und vorher muss ich noch tanken“. Er bekommt dann eine Routenplanung zum Restaurant, die auch noch einen Tankstopp einplant (vgl. Abb. 10).

Abb. 10: Kohärenter Navigationsdialog durch die Kombination mehrerer Webdienste



6 SEMANTISCHE VIDEOSUCHE UND PEER-TO-PEER-SUCHMASCHINEN

Das Aufkommen von digitalen Videokameras, Online-Videoportalen wie YouTube sowie digitaler Fernsehempfang über DVB-T ermöglicht es Anwendern, schnell große Mengen digitalen Videomaterials zu erzeugen, zu speichern und im Web zu veröffentlichen. Im Projekt InFiRe¹⁰ entwickelten Forscher des Deutschen Forschungszentrums für Künstliche Intelligenz (DFKI) unter Leitung von Thomas Breuel ein System für die inhaltsbasierte Suche in großen Videodatenbanken. Die Suche von InFiRe erfolgt dabei nach visuellen Eigenschaften des Videomaterials. Mit der Auswahl eines Keyframes aus der Videodatenbank wird die Suche nach ähnlichen Videos gestartet (vgl. Abb.11).

¹⁰ Ulges/Schulze/Keysers/Breuel 2007.

Abb. 11: Semantische Videosuche mit InFiRe

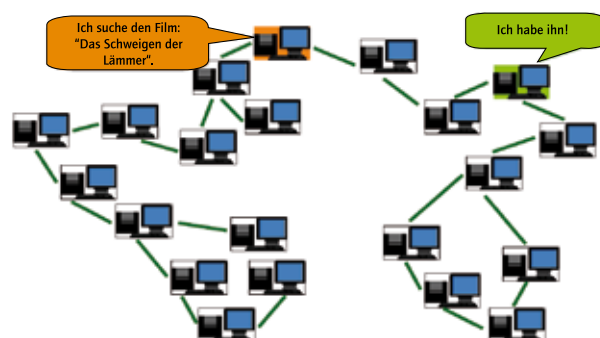


Der Vergleich des Datenbankinhalts zur Anfrage erfolgt über eine Reihe von visuellen Merkmalen wie Farbe, Textur oder Bewegung. Durch Variation der verwendeten Merkmale und deren Gewichtung zueinander kann der Anwender Einfluss auf die Eigenschaften des Suchergebnisses nehmen. Eine Gewichtung einzelner Suchergebnisse hinsichtlich ihrer Relevanz erlaubt eine weitere Verbesserung des Resultats. Eine Demonstration kann unter <http://demos.iupr.org/cgi-bin/infire.cgi> aufgerufen werden. Bei der Verwendung spezieller Videomerkmale und Distanzmaße lässt sich auch eine explizite Suche nach identischen Videoinhalten durchführen, was zum Beispiel zum Auffinden von urheberrechtlich geschütztem Material in Videodatenbanken genutzt werden kann. Momentan liegt der Schwerpunkt der Entwicklung auf der automatischen ontologischen Annotation von Videodaten aus Online-Videoportalen. Das System lernt dabei die visuellen Eigenschaften der Videos zu einem Schlagwort (zum Beispiel „Fußball“, „Schwimmen“, „Interview“, „Talkshow“) autonom, indem es automatisch Videos aus den Online-Portalen herunterlädt und deren Merkmale lernt. Die Annotationen können dann

automatisch einem Benutzer zur Auswahl vorgeschlagen werden (<http://demos.iupr.org/cgi-bin/videtagging.cgi>). In Kombination mit anderen Merkmalen können sie auch wiederum für das semantische Video-Retrieval eingesetzt werden.

Die Peer-to-Peer-Technologie hat das Potenzial, das Monopol der Webindexierung durch heutige Suchmaschinen-Firmen zu brechen. Dabei wird eine benutzerdedizierte, semantische und kontextsensitive Suchmaschine auf jedem Peer-Rechner installiert und soziale Netze werden bei der Kollaboration der Peers systematisch genutzt (vgl. Abb. 12). Über verteilte Hash-Tabellen entstehen skalierbare und selbstorganisierende Datenstrukturen und Suchverfahren über Millionen von Peers.

Abb. 12: Semantische Peer-to-Peer-Suche



Dabei werden lokal berechnete PageRank-Werte für die Relevanz von Dokumenten aggregiert und bei zufälligen Rendezvous mit anderen Peers ausgetauscht.¹¹ Es entsteht ein dezentraler Index mit lokal berechneten Autoritätswerten für die Güte eines Dokumentes, der gegen die global berechneten Werten konventioneller Suchmaschinen konvergiert, ohne dabei den hohen Aufwand zentraler Datenzentren zu erfordern.

7 KONKLUSION

In Zukunft geht die Entwicklung von Suchmaschinen hin zu Antwortmaschinen. Semantische Technologien auf der Basis maschinenverstehbarer ontologischer Annotationen von Webinhalten sind die Basis für hochpräzise Suchagenten und mobile Antwortmaschinen. Das zukünftige Web 3.0 wird aus einer Kombination der Ansätze des semantischen Webs und des Web 2.0 entstehen. Dabei wird das Prinzip der Benutzerpartizipation des Web 2.0 mit der ontologisch fundierten Beschreibung semantischer Webinhalte kombiniert. Komplexe Anfragen können nur durch die automatische Komposition semantischer Webdienste beantwortet werden, sodass ein Internet der Dienste entstehen muss. Die Semantische Videosuche wird verstärkt auf maschinellen Lernverfahren beruhen, die

¹¹ Bender/Michel/Triantafillou/Weikum/Zimmer 2006.

autonom eine ontologische Annotation aus Bildmerkmalen ableiten. Durch innovative Peer-to-Peer-Ansätze könnte langfristig das heutige Indexierungsmonopol der Suchmaschinen-Firmen gebrochen werden und durch die Nutzung von Metawissen über die jeweilige Expertise der Peer-Nutzer eine weitere Steigerung der Antwortpräzision erreicht werden.

8 DANKSAGUNG

Als Gesamtprojektleiter danke ich dem Bundesministerium für Bildung und Forschung (BMBF) für die Förderung von SmartWeb von 2004 bis 2007 unter dem Förderkennzeichen O1 IMD O1 sowie allen beteiligten Industriepartnern für die zusätzlich eingesetzten Investitionsmittel. Insgesamt ist SmartWeb eine Teamleistung von über 100 Wissenschaftlern und Softwareingenieuren, die in einer verteilten Entwicklungsumgebung ein multimodales Dialogsystem geschaffen haben, das trotz seiner weltweit einzigartigen Leistungsfähigkeit eine große Robustheit unter Realzeitbedingungen aufweist. Daher möchte ich mich an dieser Stelle bei allen Mitarbeitern des SmartWeb-Konsortiums sehr herzlich für die exzellente Kooperation bedanken.

9 LITERATUR

Bender/Michel/Triantafillou/Weikum/Zimmer 2006

Bender, Matthias/Michel, Sebastian/Triantafillou, Peter/Weikum, Gerhard/Zimmer, Christian: „P2P Content Search: Give the Web Back to the People.“ In: *5th International Workshop on Peer-to-Peer Systems (IPTPS 2006)*. Santa Barbara/USA, 2006.

Cramer/Leidner/Klakow 2006

Cramer, Irene/Leidner, Jochen/Klakow, Dietrich: „Building an Evaluation Corpus for German Question Answering by Harvesting Wikipedia.“ In: *Proceedings of LREC 2006, International Conference on Language Resources and Evaluation*. Genova/Italy, Mai 2006.

Fensel/Hendler/Lieberman/Wahlster 2003

Fensel, Dieter/Hendler, Jim/Lieberman, Henry/Wahlster, Wolfgang: *Spinning the Semantic Web. Bringing the World Wide Web to Its Full Potential*. Cambridge: MIT Press 2003.

Kaiser/Mögele/Schiel 2006

Kaiser, Moritz/Mögele, Hans/Schiel, Florian: „Bikers Accessing the Web: The SmartWeb Motorbike Corpus.“ In: *Proceedings of the LREC 2006, International Conference on Language Resources and Evaluation*. Genova/Italy, Juni 2006.

Maybury/Wahlster 1998

Maybury, Mark/Wahlster, Wolfgang (Hrsg.): *Readings in Intelligent User Interfaces*. San Francisco: Morgan Kaufmann, 1998.

Oberle et al. 2007

Oberle, Daniel/Ankolekar, Anuypriya/Hitzler, Pascal/Cimiano, Philipp/Sintek, Michael/Kiesel, Malte/Mougouie, Babak/Vembu, Shankar/Baumann, Stephan/Romanelli, Massimo/Buitelaar, Paul/Engel, Ralf/Sonntag, Daniel/Reithinger, Norbert/Loos, Berenike/Porzel, Robert/Zorn, Hans-Peter/Micelli, Vanessa/Schmidt, Christian/Weiten, Moritz/Burkhardt, Felix/Zhou, Jianshen: „DOLCE ergo SUMO: On Foundational and Domain Models in SWIntO (SmartWeb Integrated Ontology).“ In: *Journal of Web Semantics* 3 (2007).

Reithinger/Herzog/Blocher 2007

Reithinger, Norbert/Herzog, Gerd/Blocher, Anselm: „SmartWeb – Mobile Broadband Access to the Semantic Web.“ In: *KI – Künstliche Intelligenz* 2 (2007), S. 30-33.

Sonntag/Engel/Herzog/Pfalzgraf/Pfleger/Romanelli/Reithinger 2007

Sonntag, Daniel/Engel, Ralf/Herzog, Gerd/Pfalzgraf, Alexander/Pfleger, Norbert/Romanelli, Massimo/Reithinger, Norbert: „Smart Web Handheld – Multimodal Interaction with Ontological Knowledge Bases and Semantic Web Services.“ In: Huang, Thomas/Nijholt, Anton/Pantic, Maja/Pentland, Alex (Hrsg.): *Artificial Intelligence for Human Computing* (Lecture Notes in Computer Science 4451). Berlin, Heidelberg: Springer, 2007, S. 272-295.

Ulges/Schulze/Keysers/Breuel 2007

Ulges, Adrean/Schulze, Christian/Keysers, Daniel/Breuel, Thomas: „Content-based Video Tagging for Online Portals.“ In: *MUSCLE Workshop 2007, Third MUSCLE/ ImageCLEF Workshop on Image and Video Retrieval Evaluation*. Budapest/Hungary, September 2007.

Wahlster 2000

Wahlster, Wolfgang (Hrsg.): *Verbmobil: Foundations of Speech-to-Speech Translation*. Berlin, Heidelberg, New York, Barcelona, Hongkong, London, Mailand, Paris, Singapur, Tokyo: Springer, 2000.

Wahlster 2003

Wahlster, Wolfgang: „Towards Symmetric Multimodality: Fusion and Fission of Speech, Gesture, and Facial Expression.“ In: Günter, A./Kruse, R./Neumann, B. (Hrsg.): *KI 2003: Advances in Artificial Intelligence. Proceedings of the 26th German Conference on Artificial Intelligence, September 2003* (LNAI 2821). Hamburg, Berlin, Heidelberg: Springer, 2003, S. 1-18.

Wahlster 2004

Wahlster, Wolfgang: „Semantisches Web.“ In: Bullinger, H.-J. (Hrsg.): *Trendbarometer Technik: Visionäre Produkte, Neue Werkstoffe, Fabriken der Zukunft*. München, Wien: Hanser, 2004, S. 62-63.

Wahlster 2006

Wahlster, Wolfgang (Hrsg.): *SmartKom: Foundations of Multimodal Dialogue Systems*. Cognitive Technologies Series. Heidelberg: Springer, 2006, S. 644 ff.

Wahlster 2008

Wahlster, Wolfgang: „SmartWeb: Ein multimodales Dialogsystem für das semantische Web.“ In: Reuse, B./Vollmar, R. (Hrsg.): *40 Jahre Informatikforschung in Deutschland*. Heidelberg, Berlin: Springer, 2008.

> WIRTSCHAFTLICHE ASPEKTE DER SUCHTECHNOLOGIE: ES GEHT UM (VIEL) MEHR ALS (NUR) SUCHE

STEFAN WESS/JASMIN FRANZ

1 WARUM SUCHEN WIR?

Die Suche nach Information bestimmt einen Großteil unseres Arbeitsalltags. Doch ist sie keine neue Erscheinung, sondern war schon immer Teil der Arbeitsprozesse in wissensintensiven Branchen. Nur ihre Intensität und die Erwartungshaltung der Nutzer an die Informationsgewinnung haben sich im Laufe der Zeit verändert. Eines der Hauptargumente für den erhöhten Suchaufwand ist die Behauptung, dass sich die Daten so rapide vermehren, dass man förmlich in ihnen ertrinkt. Schauen wir uns das einmal genauer an: Ein White Paper der IDC (März 2007), das sich mit der Analyse des weltweiten Informationswachstums bis 2010 auseinandersetzt, macht tatsächlich erschreckende Beobachtungen und gibt eine nicht minder besorgniserregende Zukunftsvision ab: „In 2007 the amount of information created will surpass, for the first time, the storage capacity available“. Den Weg dorthin pflastern Daten, die den etwa dreimillionenfachen Umfang aller bisher geschriebenen Bücher ausmachen: „In 2006, the amount of digital information created, captured, and replicated was 1,288 x 10¹⁸ bits. In computer parlance, that's 161 exabytes or 161 billion gigabytes.“ In den nächsten zwei Jahren werden voraussichtlich mehr Daten generiert als in der gesamten bisherigen Geschichte. Und für 2010 wird das Sechsfache heutiger Datenmengen (988 Exabyte oder Bücherstapel von der Erde bis zum Planeten Pluto und zurück) prophezeit.¹

Da die menschliche Informationsverarbeitung naturgemäß limitiert ist, wird die Unübersichtlichkeit der Informationen aufgrund ihrer Masse, ihrer Granularität und der Art des Zugangs zunehmend zum Problem. Wir „dürsten“ nach Wissen, „ertrinken“ aber in Information. Wir können Relevantes von Irrelevantem nicht unterscheiden. Wir verpassen wichtige Neuerungen, weil sie „im Rauschen“ oder „in der Flut“ untergehen. Ist Suchen allein wirklich die Lösung? Wollen wir nicht lieber finden oder sogar gefunden werden? Wie finden Sie die Vorstellung, dass Sie zu jeder Zeit, an jedem Ort genau die Information erhalten, die Sie gerade brauchen – ohne großen Aufwand? „Organisation“ ist das Schlüsselwort. Empolis – als Informationslogistiker – ist der Auffassung, dass Suchen nur Ausdruck mangelnder Organisation ist, und damit sind wir nicht allein: „Lack of organisation of information is in fact the number one problem in the

¹ IDC 2007.

opinion of business professionals“, bestätigte schon im Jahr 2002 die Analytenfirma Delphi Group.² Ein heutiges reales Unternehmen besitzt etwa 70-80 Prozent unstrukturierter Daten. Neben 6 Terabyte strukturierter Daten (zum Beispiel SAP-Systeme) geht es um insgesamt 24 Terabyte Daten (die amerikanische Nationalbibliothek hat übrigens 20 Terabyte Daten), die nicht ohne Weiteres erschließbar sind, jedoch die eigentliche Wertschöpfungsquelle der Unternehmen darstellen. Eine IDC-Studie vom Oktober 2006 errechnete Kosten für die Informationsbeschaffung von 28.000 Dollar pro Jahr und Mitarbeiter, der dafür etwa 40 Prozent seiner Arbeitszeit aufbringen muss.

Es ist müßig, daraus zu schließen, dass sich die IT-Infrastruktur wirtschaftlich operierender – großer und kleiner – Firmen und Organisationen immer weniger den bereits existierenden, hoch entwickelten Technologien für das Informationsmanagement, die Datensicherheit, den Datenzugang und die Bewahrung von Informationen verschließen dürfen, wollen sie denn das Wachstum und die Heterogenität der Daten effektiv in den Griff bekommen. Ziel der IT muss es somit sein, durch semantische Methoden die Wissensarbeiter effizient zu unterstützen. Bill Gates hat dies in einer Rede im Februar 2005 auf den Punkt gebracht: „The software challenges that lie ahead are less about getting access to the information people need and more about making sense of the information they have – giving them the ability to focus, prioritize and apply their expertise, visualize and understand key data, and reduce the amount of time they spend dealing with the complexity of an information-rich environment.“

2 WIE SUCHEN WIR HEUTE?

Schaut man sich im Unternehmensalltag genauer um, erscheint die Realisierung von Bill Gates' Vision Lichtjahre entfernt. Die Suchwerkzeuge, die Wissensarbeiter im unstrukturierten Datenlabyrinth einsetzen, muten geradezu steinzeitlich an. Sie suchen meist per Stichwort im Volltext. Wird auch externer Inhalt einbezogen, greifen sie meist zum beliebten Suchmedium „Google“, das ihnen den Zugang zum weltweiten Netz (wobei das „Deep Web“ jedoch nicht berücksichtigt wird) wiederum per Suchmaske zur Eingabe von Stichworten bietet. Wir alle wissen, was bei solch einem Vorgang passiert. Der Suchende begnügt sich meist mit einem Stichwort und erhält dann entweder überhaupt keinen Treffer oder aber gleich Tausende, aufgelistet auf unzähligen Seiten in einer für den Nutzer nicht unbedingt nachvollziehbaren Reihenfolge. Spätestens hier beginnt nun der zeitaufwändige Teil der Informationsbeschaffung. Der Anwender ist gezwungen, die ihm relevant erscheinenden Hits anzuklicken, um sich dann auf der angezeigten Seite umständlich die nötigen Informationen manuell zusammenzusuchen, soweit sie überhaupt vorhanden sind. Und das mehrmals täglich.

Als ein etwas effizienterer Weg bei der Internetsuche erscheint der, sich der Dienste „vertikaler“ Suchmaschinen zuzuwenden, die im Unterschied zur „horizontalen“ Suche ein bestimmtes Themengebiet fokussieren oder eine spezifische Zielgruppe bedienen. Sucht

² Delphi Group 2002.

der Fachanwender zum Beispiel die neuesten internationalen Forschungsergebnisse auf dem Gebiet der Dentalkeramik, erhält er bei Google nach Eingabe von „dental ceramics“ 1.230.000 (!) Hits. Gibt er dagegen „ceramics“ beispielsweise auf der vertikalen Suchmaschine „Dentalproducts.net“ ein, verringert sich die Ergebnismenge auf 99 Treffer. Der Vorteil für den Nutzer besteht hier in der überschaubareren Ergebnismenge. Im Information Retrieval spricht man hierbei vom „Maß der Genauigkeit“ („Precision“ vs. „Recall“). Der „Recall“ beschreibt die Vollständigkeit eines Suchergebnisses. Er ist definiert als der Anteil der bei einer Suche gefundenen relevanten Dokumente (bzw. Datensätze) an den relevanten Dokumenten der Grundgesamtheit. Die „Precision“ beschreibt die Genauigkeit eines Suchergebnisses. Sie ist definiert als der Anteil der gefundenen relevanten Dokumente an allen bei der Suche gefundenen Dokumenten. Für die Evaluierung eines Information-Retrieval-Systems sollten die beiden zusammenhängenden Maße gemeinsam betrachtet werden. Im Allgemeinen sinkt mit steigendem Recall (mehr Treffer) die Precision (mehr irrelevante Ergebnisse) und umgekehrt sinkt mit steigender Precision (weniger irrelevante Ergebnisse) der Recall (mehr relevante Dokumente, die nicht gefunden werden), das heißt es besteht eine negative Korrelation. Stellt man das Verhältnis zwischen Recall und Precision in einem Diagramm dar, so wird der (höchste) Wert im Diagramm, an dem der Precision-Wert gleich dem Recall-Wert ist – also der Schnittpunkt des Precision-Recall-Diagramms mit der Identitätsfunktion – der „Precision-Recall-Break-even-Punkt“ genannt. Der Einsatz der vertikalen Suche ist somit nicht frei von Zweifeln über die tatsächliche Vollständigkeit relevanter Dokumente. Jedoch verfügen vertikale Suchmaschinen im Gegensatz zu horizontalen über die Möglichkeit, semantisch basiert, das heißt auf der Grundlage von spezifischen Wissensmodellen, zu operieren, was ihre Suchergebnisse tatsächlich um ein Vielfaches präziser macht.

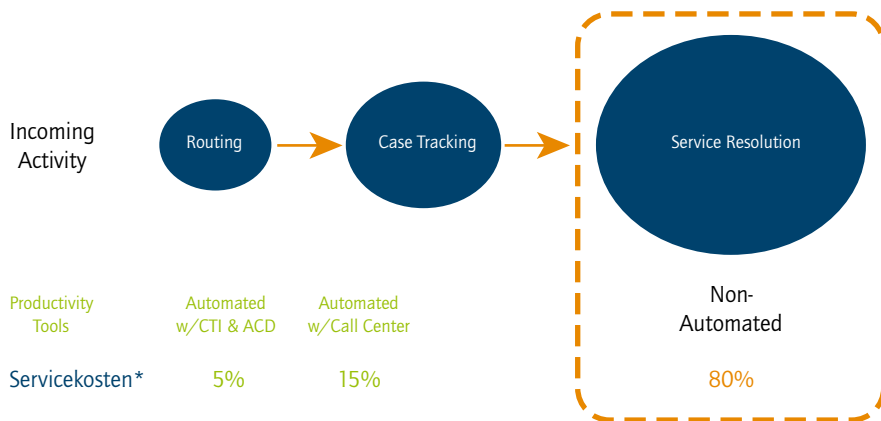
Abb. 1: Beispiel einer vertikalen Suchmaschine mit Parametrisierung

The screenshot shows the GLOBALSPEC website interface. At the top, there is a navigation bar with the GLOBALSPEC logo and a search bar. Below the search bar, there is a breadcrumb trail: "You are here: GlobalSpec.com > Electrical and Electronic Components > Power Supplies and Conditions > AC Power Sources". The main content area is titled "Search by specifications for: AC Power Sources". Below this, there is a section for "General Characteristics criteria:" with two columns of checkboxes. The first column includes "Waveform Output:" with options for "No Preference" (checked), "Sine Wave", and "Square Wave". The second column includes "Style:" with options for "No Preference" (checked), "Tooth Chamber Mount", "ATX", "Cabinet", "External Power Supply", "Desktop", and "DIN Rail". To the right of these options, there are additional checkboxes for "Triangular Wave" and "Other". On the far right, there is a "PRODUCT ARRANGEMENTS" section with a small image of a power supply unit and the text "Programmable AC Power Source, Triathlon II Generator Corp." and "Power Distribution Isolation & Management Unit".

Die anmeldepflichtige vertikale Suchmaschine „Globalspec“ ermöglicht insbesondere Experten durch eine breit angelegte Parametrisierung des gesuchten Objekts einen schnellen Zugriff auf die gewünschte Information. Je spezifischer die Anfrage, desto exakter der Treffer. Und vor allem im Endverbraucherbereich findet man auch dialoggestützte Beratungssysteme, die den Nutzer durch Fragen zum gewünschten Produkt leiten, zum Beispiel der Waschmaschinenberater auf dem Onlineportal der Firma Neckermann.

Diese Tools sind schon recht hilfreich, aber leider nicht ausreichend, wenn es um die zeitkritische Beschaffung komplexer Informationen zur Lösung von Problemen geht. Dieses Szenario bestimmt beispielsweise die Tagesarbeit im Servicebereich. Laut der SSPA Support Benchmark Study, die mittels Befragung von über 200 Technologiefirmen Schlüsseldaten zu Service & Support, Vertrieb und Marketing, Finanzen und Kundenzufriedenheit liefert, liegt die wahre Wertschöpfung im bisher nicht automatisierten Bereich der eigentlichen Erbringung der Dienstleistung, wie zum Beispiel der Lösung eines Problems. Umgerechnet auf die jeweiligen Servicekosten ergeben sich für die eigentliche Problemlösung 80 Prozent der Gesamtkosten, während auf das Routing nur 5 Prozent und auf das Case Tracking 15 Prozent (beides automatisierte Prozesse) entfallen.

Abb. 2: Ergebnisse der SSPA Support Benchmark Studie



Source SSPA Support Benchmark Study


Die Studie macht deutlich, dass das bisherige Informationsmanagement mit unstrukturierten Daten in diesem Sektor sehr kostenintensiv und stark verbesserungswürdig ist. Der Schlüssel liegt in der Überwindung der semantischen Lücke, die zwischen dem Umgang mit strukturierten und unstrukturierten Daten besteht. Im folgenden Abschnitt möchte ich darlegen, welche Technologien heute dafür zur Verfügung stehen und als Lösung geeignet erscheinen.

3 WIE WIR FINDEN

Wenn eine Stichwort- oder Volltextsuche nicht ausreichend ist, wie können wir dann besser etwas finden? Der Schlüssel hierzu liegt in der Fähigkeit zur Interpretation von Anfrage und Daten. Für eine Suchmaschine stellt sich das Internet ähnlich dar, wie der Diskus von Phaistos³ für den naiven Betrachter: Wir sehen Folgen von Zeichen in Gruppen unterschiedlicher Länge, deren Bedeutung wir nicht kennen. Tatsächlich sind Inhalt und Bedeutung der Scheibe bis heute unbekannt.

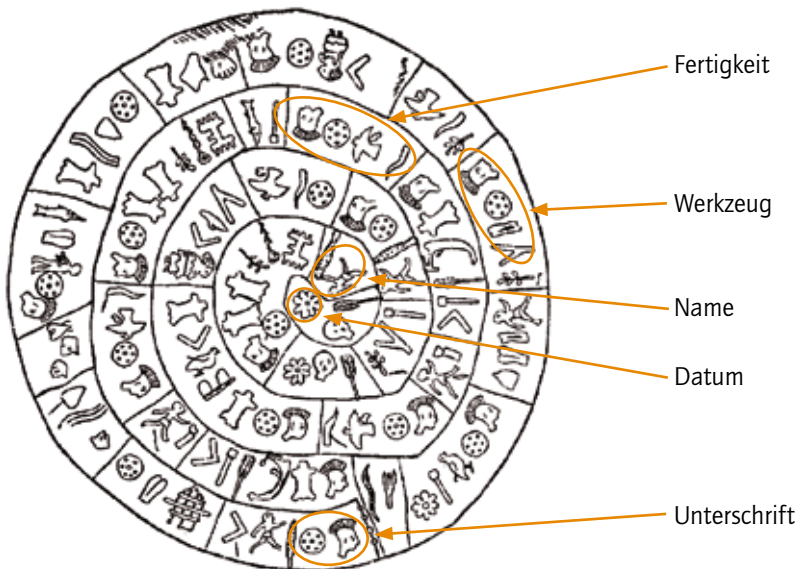
Abb. 3: Diskus von Phaistos



Eine Entschlüsselung und damit eine Interpretation der Zeichenkombinationen auf der Scheibe werden erst möglich, wenn ihre jeweilige Bedeutung bekannt ist. Dann ist es auch möglich, nach Dokumenten zu fragen, die von  unterzeichnet wurden. Diese Information über die Bedeutung wird an die Daten annotiert und diese Metadaten können wir für die Interpretation nutzen. Dabei setzen wir Wissen ein, zum Beispiel, dass ein bestimmtes Werkzeug eine bestimmte Fertigkeit erfordert.

³ Unter http://de.wikipedia.org/wiki/Diskos_von_Phaistos findet sich eine Beschreibung.

Abb. 4: Diskus von Phaistos mit (fiktiven) Metadaten



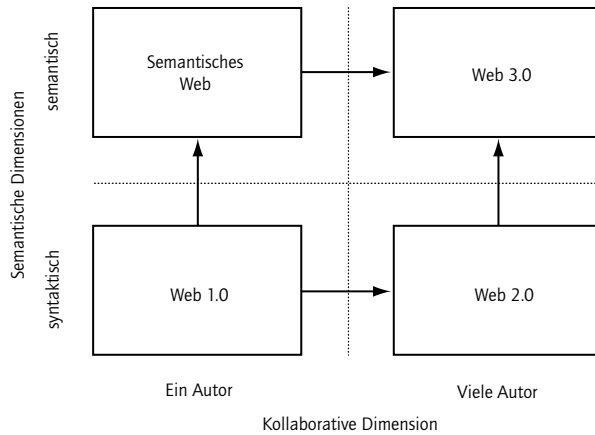
In gleicher Weise brauchen wir Suchmaschinen, die in der Lage sind, auf der Basis von Domänenmodellen, das heißt kodifiziertem Fachwissen, unstrukturierte Daten so zu interpretieren, dass wir strukturiert suchen können. „Suchen“ bedeutet in diesem Sinne die Einbeziehung des *Kontexts des Informationsbedürfnisses des Anwenders* und die Interpretation der Information in ihrer *Bedeutung in diesem Kontext*. Das Ergebnis sind die Informationen, die relevant sind für die *Erledigung der Aufgabe* des Anwenders. Dies markiert den Übergang vom rein syntaktischen zum semantischen Web.

Vertikale Suchmaschinen wie beispielsweise www.kunststoffe.de gehen diesen Weg. Sie strukturieren ihre Inhalte durch Kategorisierungen in taxonomischen Hierarchien. Darüber hinaus werden Inhalte teilweise auch annotiert, um so eine direkte strukturierte Suche zu ermöglichen. Die Navigation orientiert sich an fachspezifischen Aufgaben, zum Beispiel dem Rohstoffeinkauf. Der Preis, den der Anwender dafür zahlt, ist die potenzielle Unvollständigkeit der Inhalte. Gleichzeitig gewinnt er aber an Präzision mit Blick auf die Qualität der Suchergebnisse und damit verbunden die Geschwindigkeit, in der die gesuchten Resultate auch gefunden werden.

4 WIE FUNKTIONIERT DAS?

Ein solches Vorgehen erfordert Verfahren zur automatischen Informationsextraktion oder die manuelle Annotation von Inhalten sowie Modelle des Fachwissens. Die damit verbundenen Kosten werden zumeist überschätzt und daher gescheut. Der Nutzen durch den Effizienzgewinn bei der Informationsrecherche rechtfertigt jedoch diese Investitionen.

Abb. 5: Web-Evolution



Die Einführung von Modellen und Metadaten spiegelt die Evolution des Webs entlang der semantischen Dimension, bei der die syntaktischen Inhalte um die Beschreibung ihrer Bedeutung angereichert werden. Daneben gibt es die kollaborative Dimension des Webs. Dokumente und Inhalte werden nicht mehr nur von einzelnen Autoren erstellt, um dann von vielen konsumiert zu werden. Durch die Techniken der sogenannten „Social Software“ kann jeder die Rolle des Autors und des Konsumenten spielen. Die Inhalte werden kollaborativ von allen Beteiligten erstellt. Überträgt man diese Idee auch in die semantische Dimension, so werden auch Metadaten und Modelle kollaborativ erstellt. Wolfgang Wahlster bezeichnet dies als das „Web 3.0“.⁴

Dieses Vorgehen lässt sich auch in die Unternehmensprozesse integrieren. Andrew McAfee hat hierfür den Begriff des Enterprise 2.0 geprägt.⁵ Dabei trägt er der Tatsache Rechnung, dass viele geschäftskritische Prozesse im Unternehmen unstrukturiert sind. Dies erfordert eine Arbeitsumgebung, die dies unterstützt. Eine solche Architektur besteht aus sechs Blöcken – kurz „SLATES“ genannt:

- S earch – Inhalte personalisiert und intelligent auffinden,
- L ink – Inhalte zu einem Informationsuniversum verknüpfen,
- A uthor – Inhalte erstellen und intelligent wieder verwenden,
- T ag – Inhalte kollektiv ordnen, sortieren, navigieren,
- E xtent – neue personalisierte Anwendungen erstellen (Mashups),
- S ignal – über Neuigkeiten automatisch informiert werden.

Unter Einbeziehung der semantischen Dimension werden so unternehmensinterne Web 3.0-Anwendungen entstehen, die die Produktivität der Wissensarbeit nachhaltig steigern.

⁴ Fensel/Wahlster/Lieberman/Hendel 2003.

⁵ Vgl. McAfee 2007.

5 PRAXISBEISPIEL „KNOX“: VERTIKALE SUCHE UND ENTERPRISE 2.0 IN DER BETRIEBLICHEN PRAXIS

Zur Exemplifizierung der Umsetzung der beschriebenen Erkenntnisse in die betriebliche Praxis soll die von empolis für Siemens IT Solutions and Services entwickelte Enterprise 2.0-Anwendung „Knox“ dienen.

Siemens IT Solutions and Services, die frühere Siemens Business Services, bietet IT-Lösungen, IT Outsourcing und Softwarelösungen für die Industrie aus 40 Ländern an. Zu ihren Kunden zählen unter anderem BASF, BBC, Barclays, DaimlerChrysler, EADS und Airbus.

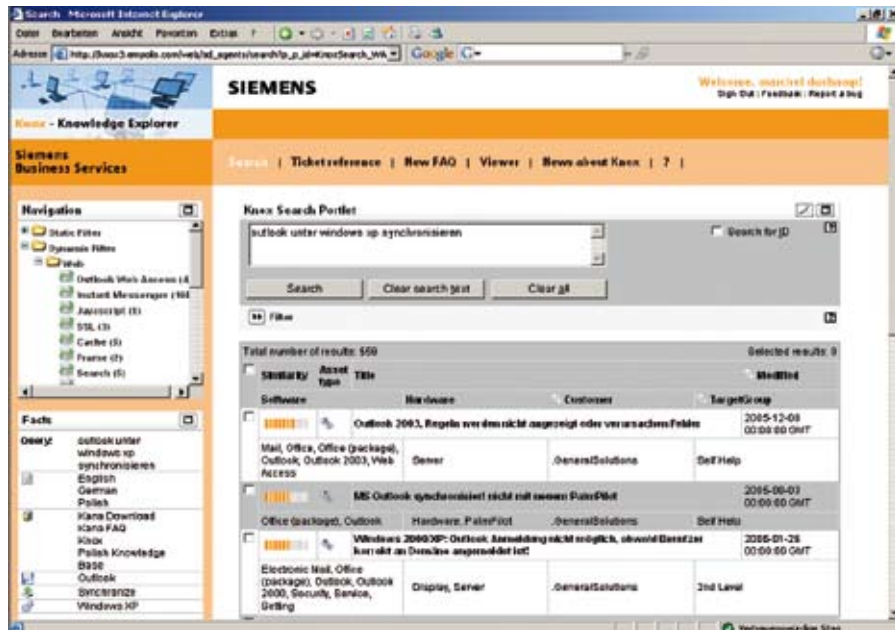
Die Global Business Unit Operation Related Services (ORS), die ihren Kunden Supportleistungen rund ums Outsourcing von Rechenzentren, Desktop Services, lokalen Netzen und Callcentern anbietet, beauftragte empolis mit der Realisierung eines web-basierten Informationsportals für ihre Mitarbeiter im Callcenter. Dieses wurde in das bestehende Call Tracking System (Remedy Action Request System) integriert, so dass problemlos Daten aus dem Tracking System an das Knowledge Management System übergeben werden können. Im Ergebnis garantiert der Knox Knowledge Explorer einen intelligenten, einheitlichen Zugriff auf den weltweit verteilten Bestand von unterschiedlichsten Serviceinformationen im Callcenter von Siemens IT Solutions and Services.

Folgende wesentliche Ziele wurden mit der Einführung dieser Wissensbasis verfolgt:

- Verkürzung der Bearbeitungszeit der Trouble-Tickets,
- Erzielung einer höheren Direktlösungsrate im First-Level Support durch einen Wissenstransfer aus dem Second-Level Support,
- weltweit einheitlicher Zugriff auf den heterogenen Informationsbestand,
- Verbesserung der Support-Qualität durch standardisierte und autorisierte Serviceinformationen.

Im Rahmen des Projekts wurde zunächst der Status Quo analysiert und anschließend auf Basis der Ergebnisse das zukünftige System spezifiziert. Der daraufhin entwickelte Knox Knowledge Explorer basiert im Kern auf der e:Service Lifecycle Suite der empolis GmbH. Mithilfe dieser Technologie können unterschiedliche Datenquellen indiziert und intelligent abgerufen werden. Durch den weltweit einheitlichen Zugriff auf Informationen werden zukünftig Redundanzen bei der Erstellung von Dokumenten vermieden. Die Antwortqualität ist unabhängig vom Standort des jeweiligen Callcenters.

Abb. 6: Praxisbeispiel Knox bei Siemens



Der Zugriff auf den Informationsbestand basiert auf einem mehrsprachigen Wissensmodell, mit dessen Hilfe eingehende Anfragen linguistisch analysiert werden, um anschließend relevante Dokumente in einer Ergebnisliste anzuzeigen. Das mehrsprachige Wissensmodell garantiert dabei Treffer in allen Sprachen, die der Benutzer versteht, auch wenn zum Beispiel nur in Deutsch angefragt wurde. Neben der natürlichsprachigen Formulierung des Problems kann der Benutzer über eine Baumstruktur einerseits direkt auf die gesuchten Dokumente zugreifen und andererseits zusätzliche Filterkriterien bei der Suche setzen. Die Baumstruktur wird aus dem zentral gepflegten Wissensmodell gespeist. Die Konzepte „Suche“ und „Navigation“ greifen somit ideal ineinander.

Die Verwendung einer wissensmodellbasierten Suchtechnologie unterscheidet Knox deutlich von den eingeschränkten Suchfunktionen der bis dahin bei ORS eingesetzten Systeme und optimiert den Bearbeitungsprozess im Callcenter. Das Wissensmodell stellt sicher, dass Knox die eingegebenen Probleme „versteht“ und präzise die Informationen anzeigt, die auf die aktuelle Problemsituation angewendet werden können. „Mit der Einführung von Knox schaffen wir ein global einsetzbares und in den Ticket-Workflow nahtlos integriertes Knowledge Tool – eine wichtige Basis für die Zufriedenheit unserer Kunden“, so Dieter Linder, Projektverantwortlicher bei Siemens IT Solutions and Services ORS.

Derzeit läuft Knox produktiv in vier Ländern und im Testbetrieb in zwei weiteren. Das Team um Linder hat Prozesse, Begriffe, Speicher und Rollen über alle teilnehmenden Organisationen etabliert. In Zahlen nutzen Knox 2000 Agenten, 50 Experten erstellen Wissens Elemente, 30 Wissensförderer trainieren und unterstützen die Agenten bei der Arbeit und fungieren als deren Vermittler zur Verwaltung. Nach Fertigstellung des Berichtssystems wird man auch genaue Kennzahlen erhalten. Es ist geplant, nach und nach weitere Länder an Knox anzuschließen. Das nächste Projekt sorgt für die Optimierung der Integration von neuen globalen CRM Ticketing Prozessen.

6 FAZIT

Der Titel dieses Beitrags hat den Schluss eigentlich schon vorweggenommen. Suche als solche ist nicht die Panazee der Informationserschließung, kann jedoch auf Basis intelligenter Technologie im Zusammenspiel mit wesentlichen Voraussetzungen wie geeigneter Datenorganisation, Modellierung der Fachdomänen, Integration in die Geschäftsprozesse und einfacher Bedienbarkeit fast Wunder wirken. Zumindest rückt das Ziel, die richtige Information zur richtigen Zeit der richtigen Person bereitzustellen, mit dem Einsatz semantischer Suchtechnologie bedeutend näher. Wenn wir nämlich über die Unternehmensgrenzen hinaus blicken und auch Produkte und Produktionsmittel berücksichtigen, dann entsteht die Vision einer umfassenden Kommunikationsinfrastruktur mit sich dynamisch anpassenden Anwendungen („Ubiquitous Computing und Ambient Intelligence“). Daraus folgt, dass wenn wir heute über ein zukünftiges Web 3.0 reden, wir letztlich die semantische Integration dreier Welten und ihrer Inhalte meinen:

- Das Internet der Dienste (Processes) – Web Services, Business Rules, service-orientierte Architekturen (SOA)
- Das Internet der Dinge (Products) – Radio Frequency Identification (RFID) und Sensornetzwerke
- Das Internet der Menschen (People) – Soziale Netzwerke

Es geht also tatsächlich um viel mehr als Suche. Das Unternehmen 2.0 wird sich daran messen lassen müssen.

7 LITERATUR

Delphi Group 2002

Delphi Group: *White Paper*. URL: <http://www.delphigroup.com/research/whitepapers/asp> [Stand: 20.09.2007].

Fensel/Wahlster/Lieberman/Hendler 2003

Fensel, Dieter/Wahlster, Wolfgang/Lieberman, Henry/Hendler, James: *Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential*. MIT Press 2003.

IDC 2007

IDC: *The Expanding Digital Universe*. URL: http://www.emc.com/about/destination/digital_universe/pdf/Expanding_Digital_Universe_IDC_WhitePaper_022507.pdf [Stand: 20.09.2007].

McAfee 2007

McAfee, Andrew: *The Impact of Information Technology (IT) on Businesses and their Leaders*. URL: http://blog.hbs.edu/faculty/amcafee/index.php/faculty_amcafee_v3/C24 [Stand: 20.09.2007].

> DIE BEDEUTUNG VON SUCHE IM ONLINE-GESCHÄFT

MICHAEL BUSCHMANN

1 EINLEITUNG

In den vergangenen Jahren hat das Thema Suche in der Wahrnehmung von Online Professionals, zunehmend aber auch der breiten Öffentlichkeit, kontinuierlich an Aufmerksamkeit gewonnen. Ein wesentlicher Treiber für diese Entwicklung war sicherlich die bemerkenswerte Entwicklung der Firma Google vom einstigen Start-up-Unternehmen hin zum dominanten Internet-Giganten innerhalb nur weniger Jahre. Mit dem Erfolg von Google wurde zunehmend auch eine Diskussion angeregt, die sich um die Machtposition von derart erfolgreichen Suchmaschinen dreht. Tatsächlich scheint erst in den vergangenen Jahren so richtig klar geworden zu sein, welche wichtige Position die Suchmaschinen innerhalb der Internet-Ökonomie einnehmen.

2 BEDEUTUNG DER SUCHE FÜR DEN INTERNET-NUTZER

Wie aus einer regelmäßig erscheinenden Studie der AGOF (Arbeitsgemeinschaft Online-Forschung e.V.) hervorgeht,¹ gehört die Suche bzw. Recherche im Internet zu den am häufigsten durchgeführten Nutzer-Aktivitäten (Abb. 1). Sie liegt aktuell mit ca. 84,2 Prozent fast gleichauf mit der Nutzung von E-Mail (85,0 Prozent). Erst mit einem relativ deutlichen Abstand folgt die Nutzung von Nachrichtenquellen im Internet (59,6 Prozent).

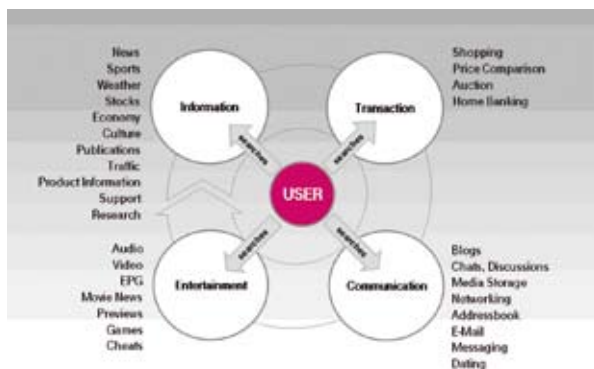
Abb. 1: Nutzeraktivitäten im Internet



¹ Vgl. AGOF 2006.

Tatsächlich ist der Vorgang der Suche im Internet nahezu allgegenwärtig. Betrachtet man die unterschiedlichen Interaktionsfelder (Information, Entertainment, Kommunikation, Transaktion, Abb. 2), so kann man innerhalb nahezu aller wesentlichen Produktgattungen die Existenz und Notwendigkeit von Suchmechanismen nachvollziehen. Die Vielfalt der uns via Internet zur Verfügung stehenden Daten, Informationen und Produkte lässt sich ganz offensichtlich nur noch mithilfe von Suchtechnologie strukturieren und damit für uns effektiv nutzbar machen.

Abb. 2: Suche im Kontext der Interaktionsfelder des Nutzers



3 BEDEUTUNG DER SUCHE FÜR ONLINE-UNTERNEHMEN

Welche Rolle spielen Suchmaschinen und Suchmaschinentechnologien nun aus der Sicht von Unternehmen, speziell Online-Unternehmen und Portalen? Suchmaschinen sind in erster Linie ein unverzichtbarer Zusammenführer von Angebot und Nachfrage. Sie unterstützen den Nutzer bei der Ermittlung von für ihn relevanten Informationen, Dienstleistungen und Produkten. Im Moment seiner Nachfrage erlauben sie den unmittelbaren Zugriff auf infrage kommende Angebote.

Da ist es nicht verwunderlich, dass viele Betreiber von Online Shops und Portalen aktive Suchmaschinenoptimierung betreiben, um sicherzustellen, dass ihr Angebot möglichst hoch in den Ergebnislisten gefunden wird. Doch bei den allermeisten kommerziellen Suchbegriffen ist die Konkurrenz inzwischen so groß, dass technische Optimierung alleine keine Garantie für eine hohe Platzierung in den natürlichen Ergebnissen einer Suchmaschine darstellt. Daher hat es nicht lange gedauert, bis auf dieser Grundlage auch ein Geschäftsmodell etabliert wurde, welches wir heute als „Suchwort-Vermarktung“ oder „Keyword Advertising“ kennen. Hier geht es schlicht und einfach darum, dass derjenige, der in einer Ergebnisliste ganz oben stehen will, dafür den höchsten Preis je Klick zu zahlen bereit sein muss. Als Portalbetreiber kann man also sowohl über

die technische Optimierung seines Angebotes, als auch über den Kauf von Keywords für einen steten Besucherstrom auf seinem Portal sorgen. Sofern es sich mit den geschäftlichen Zielen vereinbaren lässt (und die Portalinhalte dafür geeignet sind), kann man jedoch auch zum Nutznießer des Geschäftsmodells werden und eine Partnerschaft beispielsweise mit Google oder Yahoo! eingehen, indem man „Sponsored Links“ in sein eigenes Portalangebot integriert. Dies kann entweder über die Einbindung von bezahlten Links im Umfeld von redaktionellen Inhalten geschehen, oder über die Integration einer Websuche (nebst Sponsored Links) im eigenen Portal. Speziell die Einbindung von kontextuellen Sponsored Links ist ein wesentlicher Bestandteil vieler Geschäftsmodelle im Zeitalter des Web 2.0.

Doch bei der aktiven Nutzung von Suche bzw. Suchtechnologie ist man nicht allein auf Google & Co. angewiesen. Der Einsatz von Suchtechnologie in der eigenen IT kann auch auf direkte Weise Geschäftsziele unterstützen. In den vergangenen Jahren hat sich innerhalb des Suchemarktes das Segment „Enterprise Search“ herausgebildet. Die führenden Anbieter in diesem Bereich heißen, sofern man aktuellen Untersuchungen von Forrester und Gartner folgt,² interessanterweise nicht Google oder Microsoft (obwohl beide Anbieter hier präsent sind), sondern Fast, Autonomy oder Endeca. Diese Firmen bieten hoch entwickelte Suchmaschinenlösungen an, die nach den gleichen technischen Prinzipien arbeiten, die auch Google anwendet, jedoch sind ihre Produkte in besonderem Maße auf die Bedürfnisse von Unternehmen eingerichtet. Sie erlauben es den Unternehmen, eigene vertikale Suchen zu entwickeln und damit die eigenen Daten und Inhalte für den internen und externen Gebrauch zu strukturieren. Die Anbieter von Enterprise Search-Lösungen finden ihre Kundschaft verstärkt im Medien- und Verlagsgeschäft, dem sie damit, so ihr Selbstverständnis, die technischen Mittel für den Kampf gegen ein übermächtiges Google liefern. Doch Suchtechnologie spielt auch eine große Rolle im Kontext von Intranets (Knowledge Management), Market Research (Crawling und Analyse von Wettbewerbsseiten, Kunden-Foren), CRM, eCommerce und vielen weiteren Geschäftsfeldern. Betrachten wir nachfolgend den Einsatz von Suche und Suchtechnologie am Beispiel der Deutschen Telekom genauer.

4 BETRIEB EINES EIGENEN SUCHPORTALS UNTER „SUCHE.T-ONLINE.DE“

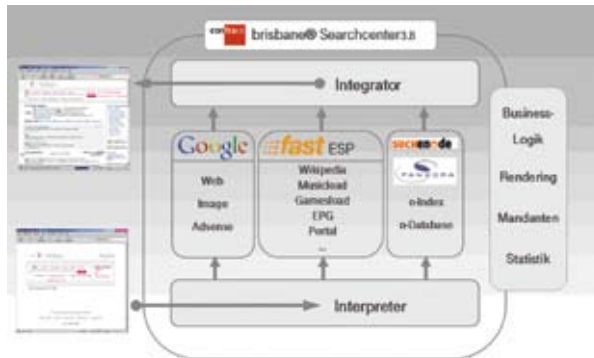
Neben Google, Yahoo und Microsoft, den weltweit führenden Anbietern für Websuche, gibt es eine Vielzahl von weniger bekannten, kleineren Suchmaschinen. Viele von diesen Suchmaschinen finden sich auf bekannten Portalen. In der Regel basieren sie auf einer Partnerschaft mit einer der großen Suchmaschinen bzw. Index-Lieferanten (Google, Yahoo!). Hier kann noch zwischen dem Bezug eines Ergebnis-Feeds für Web- und Bildersuche („Natural Search“) einerseits und dem Bezug eines Feeds mit bezahlten Ergebnissen („Paid Search“) andererseits differenziert werden. Relevante Anbieter für Natural Search sind Google, Yahoo! und Ask.com. Diese Unternehmen betreiben einen eigenen weltweiten Web-Index. Paid Search-Resultate kommen meist entweder von Google oder

² Forrester Wave 2006, Gartner Magic Quadrant 2006.

Yahoo! Search Marketing, die beide über einen sehr großen Kundenstamm auf Seiten der Werbetreibenden verfügen. Google und Yahoo! teilen die aus den Sponsored Links resultierenden Umsätze mit den Partnerportalen. Das sogenannte „Revenue-Sharing“ beginnt für kleinere Partner bei ca. 50 Prozent und steigt in Relation zum Volumen der Suchanfragen bzw. nach Verhandlungsgeschick weiter an.

T-Online kooperiert bei der Web- und Bildersuche seit mehreren Jahren erfolgreich mit Google. Vom Nutzer angefragte Begriffe werden an Google übermittelt. T-Online erhält darauf als Antwort einen XML-Feed mit den entsprechenden Ergebnissen. Die Darstellung und die zugrunde liegende Business-Logik liegen jedoch allein in der Verantwortung von T-Online. Das Frontend bzw. die Suchergebnisseite wird daher über eine eigene Suche-Plattform abgebildet (Abb. 3).

Abb. 3: Suche-Plattform von T-Online



T-Online nutzt die Websuche aktiv, um Kundenwünsche bzw. Nachfrage mit eigenen Angeboten und Inhalten zusammenzuführen. Zu diesem Zweck verwaltet die T-Online-Sucheplattform mehr als 26.000 unterschiedliche Keywords, die wiederum auf ca. 600 unterschiedliche Ergebnissets verlinken, welche mit den unter www.t-online.de offerierten Services, Produkten und Inhalten korrespondieren. Sofern der Nutzer einen Begriff, zum Beispiel „Bundesliga“, eingibt, und dieser mit einem passenden Angebot bei T-Online gematcht werden kann, wird noch oberhalb der Sponsored Links ein kleiner Ergebnisbereich mit T-Online-Treffern angezeigt. Hier findet der Nutzer nun den direkten Zugang zu Bundesliga-Videos, Tabellen und Spielberichten.

Damit hat T-Online für die Suchmaschine eine gute Balance zwischen Unterstützung eigener Produkte und der Möglichkeit zur Monetarisierung über Sponsored Links gefunden. Zusätzlich wird so aber auch erreicht, dass T-Online-Nutzer durch das Such-

portal einen T-Online-spezifischen Mehrwert erfahren. Denn auch Suchbegriffe wie „POP3“, „Hilfe“, „Sicherheit“, „WLAN-Router“, „Eumex“ oder „DSL“ führen den Nutzer zuverlässig und schnell zu Informationen und Support rund um unsere Produkte.

Die Websuche ist innerhalb des T-Online Suchportals aber nur ein Bestandteil, wenn auch ein viel genutzter. T-Online bietet zusätzlich zahlreiche weitere Suchbereiche an, unter anderem Lokale Suche, Shopping, Musik, Filme und Homepages. Ein Teil dieser Suchen basiert wiederum auf anderen Partner-Feeds. So liefert suchen.de (ein Angebot der t-info GmbH) die Lokale Suche, während die Shopping-Resultate von Pangora Lycos, einem führenden Anbieter in diesem Segment, zugeliefert werden. Nahezu alle weiteren Suchbereiche werden jedoch von T-Online selbst produziert und basieren auf der oben gezeigten technischen Plattform. Neben der obligatorischen Portalsuche, die dem Nutzer das Finden redaktioneller Inhalte unter www.t-online.de ermöglicht, bietet das T-Online-Netzwerk eine Vielzahl weiterer Portale, die für eine Erschließung durch Suchtechnologie gut geeignet sind, wie Softwareload, Gamesload, AutoScout24, Jobscout24, um nur einige zu nennen. Die Integration solcher Inhalte erlaubt uns eine klare Differenzierung gegenüber anderen Suchportalen bei gleichzeitiger Fokussierung auf unsere Kunden.

Bei T-Online begreifen wir das Suchportal in erster Linie als einen alternativen Zugang zu unseren eigenen Inhalten, welcher die klassische Portal-Navigation ergänzt. Ein weiterer wichtiger Aspekt mit Bezug auf unser eigenes Angebot ist, dass wir mit durchschnittlich ca. 2 Mio. Suchanfragen pro Tag permanent „das Ohr am Kunden haben“. Tatsächlich liefert die Suche uns täglich neue und wichtige Erkenntnisse über die Bedürfnisse unserer Nutzer, vor allem in Bezug auf unsere Produkte. Die Suche ist für uns somit auch ein wichtiges Marktforschungs-Tool. Ein Ziel für die nahe Zukunft ist es daher, diese Daten besser zu erschließen und den Redakteuren und Produktmanagern von T-Online die für sie relevanten Informationen, Trends und Analysen möglichst in Echtzeit zur Verfügung zu stellen.

Seitdem T-Online im Februar 2006 ein Redesign der Suche durchgeführt hat, welches vor allem eine stärkere Ausrichtung des Interface-Designs an dem von Google, Yahoo! und anderen populären Suchen beinhaltet, konnte die Anzahl der Suchanfragen nahezu verdoppelt werden. Die Traffic-Zuführung aus der Suchmaschine in die eigenen Portale (www.t-online.de, Gamesload, Musicload, etc.) wurde mithilfe der Produktergebnisse vervierfacht. Diese Entwicklung bestätigt, dass T-Online mit seiner Definition des Suchportals auf einem recht guten Weg ist. Gleichwohl ist dieser Erfolg natürlich angesichts der dominanten Vormachtstellung von Google (ca. 89,3 Prozent Marktanteil) zu relativieren. Die T-Online-Suche liegt mit ca. 2,1 Prozent Marktanteil (Anfang 2006 noch 1,2 Prozent) aktuell knapp vor Microsoft (aktuell ca. 1,6 Prozent) und hinter Google und Yahoo (ca. 3,0 Prozent) auf Platz 3.³

³ Webhits 2007.

5 EINBINDUNG VON SPONSORED LINKS IM KONTEXT VON REDAKTIONELLEN ARTIKELN

Diese Werbeform, von Google „AdSense for Content“ genannt, wird von T-Online ebenfalls genutzt. Dabei werden innerhalb der Portale spezielle Flächen definiert, auf denen später von Google kontextuelle Sponsored Links angezeigt werden. Zu diesem Zweck durchsucht Google die betreffenden Seiten des Angebotes in regelmäßigen, vergleichsweise kurzen Abständen und extrahiert aus den Artikeln oder Texten die wesentlichen Keywords, zu denen dann im Folgeschritt passende Anzeigen ausgeliefert werden. Als sogenanntes „General-Interest-Portal“ kann T-Online mit zahlreichen Themen aufwarten, in deren Kontext sich diese Werbeform hervorragend einbinden lässt. Letztlich ist es eine Frage der finanziellen Performance, in welchen Teilen des Portals eine Einbindung von AdSense for Content stärker oder weniger stark praktiziert wird. In jedem Fall bietet AdSense for Content speziell bei sehr spezifischen Inhalten, für die möglicherweise kein passender Werbe- oder Sponsoringpartner gewonnen werden kann, eine gute Möglichkeit zur Refinanzierung.

6 SEARCH ENGINE MARKETING (SEM)

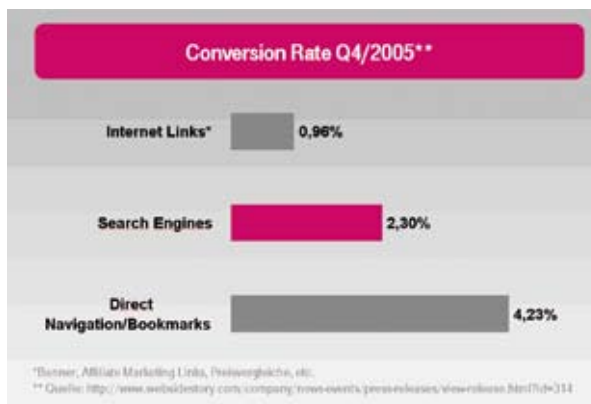
„Suchmaschinenmarketing“ (SEM) ist der Oberbegriff für die Platzierung von kommerziellen Ergebnissen (Sponsored Links) in Suchergebnissen (bei Google: AdSense for Search) und Content-Angeboten (AdSense for Content) aus der Sicht des Werbetreibenden. In der Regel buchen spezielle SEM-Agenturen zu diesem Zweck für den Werbetreibenden relevante Keywords und Platzierungen innerhalb einer Suchergebnisseite. Je höher die Platzierung und damit die Visibilität ausfällt, desto höher ist der Preis, der für einen Klick zu entrichten ist. Der Werbetreibende bleibt so lange in der Ergebnisseite des von ihm gebuchten Keywords platziert, bis sein Budget aufgebraucht ist. Anders als bei vielen anderen Werbemodellen wird hier also immer nur für eine konkret erbrachte Leistung (Klicks = Traffic-Zuführung) gezahlt. Bei Abschluss der Buchung ist damit prinzipiell auch klar, welche Leistung in Form von Klicks erbracht werden wird bzw. welcher Traffic auf der Zielseite erwartet werden kann. Damit ist diese Werbeform für den Werbetreibenden besonders gut kalkulierbar, was die außerordentliche Beliebtheit und den Erfolg im Markt erklärt.

Abb. 4: Online-Werbemarkt



Tatsächlich nimmt das Keyword Advertising laut OVK-Prognose im Jahr 2007 mit ca. 1,148 Mrd. Euro bereits ca. 46 Prozent des gesamten Online-Werbemarktes (2,533 Mrd. Euro) ein (Abb. 4).⁴ Diesen Erfolg kann man gut nachvollziehen, wenn man betrachtet, wie effizient Keyword Advertising gegenüber den zahlreichen klassischen Alternativen ist. Anfang 2006 veröffentlichte das US-Unternehmen Websidestory, ein Anbieter von Portalanalyse-Software, eine Studie, die aufzeigte, dass Traffic aus Suchmaschinen bei kommerziellen Webangeboten mit einem Anteil von 2,3 Prozent deutlich stärker zu Konversion (Transaktion, Abverkauf, Download) führt als andere Verlinkungen (Banner, Affiliate Marketing, Preisvergleiche), die nur eine durchschnittliche Konversion von ca. 0,96 Prozent aufweisen (Abb. 5).

Abb. 5: Suchmaschinen liefern qualifizierten Traffic



⁴ OVK 2007.

Insofern ist also der aus Suchmaschinen stammende Traffic von einer besonderen Qualität und somit von Bedeutung für den Erfolg eines kommerziellen Internet-Angebotes, was gleichzeitig auch die große Abhängigkeit von Suchmaschinen dokumentiert. Bei T-Online wird dieser Traffic als „qualifizierter Traffic“ bezeichnet. Von diesem Traffic kann man eigentlich nie genug haben, da er stets auf der unmittelbaren Willensbekundung des Nutzers basiert, Ergebnisse zu einem speziellen Suchbegriff zu finden. Daher gilt es, permanent zu prüfen, wie man den qualifizierten Traffic steigern kann.

7 SEARCH ENGINE OPTIMIZATION (SEO)

Die zweite Möglichkeit der Platzierung innerhalb eines Suchergebnisses ist die Optimierung des eigenen Angebotes hinsichtlich der Platzierung innerhalb des natürlichen Rankings einer Suchmaschine – also der nicht bezahlten Ergebnisse. Diesen Vorgang nennt man „Suchmaschinenoptimierung“ (meist englisch mit „SEO“ abgekürzt). Da der Wettbewerb groß ist und die Suchmaschinenbetreiber von Zeit zu Zeit die Arbeitsweise ihrer Suchalgorithmen ändern (um dem sogenannten Suchmaschinen-Spamming vorzubeugen), muss man diese Aktivität in regelmäßigen Abständen wiederholen. Grundlage für die SEO-Optimierung einer Webseite ist zunächst eine Keywordanalyse. In der Keywordanalyse wird ermittelt, welche Begriffe und Wort-Kombinationen der Nutzer in die Suchmaschine eingibt, um zum Beispiel ein Produkt oder eine spezielle Dienstleistung zu finden. Bei T-Online können wir zu diesem Zweck glücklicherweise auf unseren Datenbestand aus der Websuche zurückgreifen. Da wir ziemlich genau wissen, was der Nutzer sucht, kennen wir auch die Keywords, auf die hin das Angebot optimiert werden muss oder zu denen wir SEM-Maßnahmen buchen wollen.

Viele Unternehmen lassen sich heute von SEO-Agenturen beraten. Diese Spezial-Dienstleister haben in den vergangenen Jahren einen regelrechten Boom erlebt. Doch nach unserer Erfahrung bieten sie oftmals nur additive Maßnahmen, die beispielsweise im Rahmen von Marketing-Kampagnen den Traffic auf spezielle Landing Pages stärken, gehen jedoch nicht immer auf die grundsätzlichen Problemstellungen ein. Dies kann durchaus sinnvoll sein, insbesondere aus Sicht von Vertriebseinheiten. Allerdings sollte man nicht versäumen, die SEO-Thematik grundsätzlich anzugehen und sicherzustellen, dass das eigene Portal SEO-Erfordernissen folgt und die am Portalbetrieb beteiligten Mitarbeiter die wesentlichen Prinzipien kennen und beachten. Dafür bedarf es zunächst eines Bewusstwerdungsprozesses im Management. Erster Schritt zu einem SEO-optimierten Portal ist grundsätzlich die Erkenntnis, dass die Themen Design/Layout, HTML-Coding/Content Management und Redaktion integriert betrachtet werden müssen. Alle drei Themenfelder haben großen Einfluss auf den Erfolg von SEO-Maßnahmen – es gibt hier starke Wechselwirkungen. In den vergangenen Jahren stand vor allem das Thema „Usability“, also die Bedienbarkeit aus Sicht des Nutzers, im Fokus. Genauso wie die Bedienbarkeit aus Sicht der Nutzer ein wesentlicher Erfolgsfaktor ist, so ist auch die Ver-

ständigkeit von Webseiten für die Suchmaschinen eine wichtige Voraussetzung. Sorgt SEO für maximalen Traffic auf ein Angebot, so beschäftigt sich Usability damit, diesen Traffic in der Folge optimal zu konvertieren. Es ist also offensichtlich, dass grundsätzlich eine Balance zwischen diesen beiden auf die Gestaltung und den Aufbau einer Website Einfluss nehmenden Faktoren anzustreben ist.

Wie lässt sich SEO nun im Rahmen des Portalbetriebs erfolgreich verankern? Zunächst einmal ist sicherzustellen, dass bei sämtlichen Neuentwicklungen bereits in der Konzeptionsphase eine erste Überprüfung auf SEO-Verträglichkeit stattfindet. Damit nimmt SEO unmittelbar Einfluss auf die Gestaltung eines Portals. Auch wenn Flash und Java aus Gründen der Gestaltung gerne als Module innerhalb von Webseiten verbaut werden, empfiehlt es sich zu prüfen, inwieweit ähnliche Darstellungen in HTML oder DHTML realisiert werden können, da nur diese beiden Formate von Suchmaschinen gelesen werden können. Ein Javaticker erlaubt der Suchmaschine keinerlei Zugriff auf die dahinter liegenden Nachrichten – für die Suchmaschine existieren sie daher nicht und können darüber auch nicht gefunden werden. Sofern das Unternehmen spezifische Entwicklungsprozesse definiert hat, macht es daher grundsätzlich Sinn, einen festen Prozessschritt für SEO zu etablieren.

Der SEO-Einfluss kann so weit reichen, dass man auch die Rubrizierung eines Portals mit Blick auf eine Stärkung des Rankings gestaltet und beispielsweise Begriffe in der Hauptnavigation verwendet, die besonders oft als Suchbegriffe verwendet werden. Auch wenn zum Beispiel „Wetter“ oder „Routenplaner“ klassischerweise Unterrubriken von „Nachrichten“ oder „Reise“ darstellen, kann es Sinn machen, diese beiden Navigationspunkte gleichberechtigt in der Hauptnavigation anzubieten. Beide Begriffe gehören zu den am stärksten nachgefragten Suchbegriffen und nehmen dadurch einen sehr positiven Einfluss auf das allgemeine Ranking einer Webseite. Aus den gleichen Überlegungen heraus sollte vermieden werden, mit spezifischen Produktnamen innerhalb von Navigationen zu arbeiten. Stattdessen sollten möglichst allgemeingültige Begriffe verwendet werden. Letztlich geht es darum, die verwendete Sprache mit den Suchgewohnheiten der Nutzer zu synchronisieren.

SEO macht auch vor der eigentlichen redaktionellen Arbeit nicht Halt. Die wiederholte Nutzung von Keywords in Teasern, Überschriften und Textbodies steigert ebenfalls die Auffindbarkeit von Inhalten. Die weit verbreitete Gewohnheit, aus einem Teaser mit „Mehr“ auf den eigentlichen Artikel zu verlinken, ist ebenfalls ein No-Go. Die Links sollten stattdessen immer Keywords beinhalten, die mit der eigentlichen Botschaft des Artikels korrespondieren.

Um sicherzustellen, dass die Redaktion die Grundprinzipien der Suchmaschinenoptimierung nachhaltig umsetzt, empfehlen sich regelmäßige Schulungen und Check-Ups für die einzelnen Portale. Tatsächlich prüfen wir einmal pro Jahr jedes Teilportal oder Vertical bei T-Online hinsichtlich seiner SEO-Konformität und entwickeln in der Folge entspre-

chende portalspezifische Empfehlungen. Grundsätzlich lassen sich Redakteure ungern in ihr Handwerk hinreinen – das liegt in der Natur des Berufsstands. Betrachtet man mit ihnen gemeinsam die wirtschaftlichen Auswirkungen, die SEO auf den Erfolg eines Portals haben kann, so kann man sie in der Regel schnell für die aktive Berücksichtigung von SEO-Prinzipien gewinnen:

Das Portal onSpiele.de war einer der Vorreiter in Sachen SEO bei T-Online. Anfang 2006 nahmen wir eine Nullmessung vor und ermittelten, dass ca. 12 Prozent des monatlichen Traffics aus Suchmaschinen zugeführt wurde. Über das Jahr wurden zahlreiche, überwiegend redaktionelle Maßnahmen umgesetzt, die dazu beitrugen, dass der Suchmaschinenanteil am Gesamt-Traffic auf ca. 18 Prozent im Dezember 2006 anstieg. Dies bedeutete einen zusätzlichen Traffic von 300.000 Page Impressions. Hätte onSpiele.de sich diesen zusätzlichen Traffic über die Schaltung von Sponsored Links zukaufen wollen, so hätten dafür im Dezember 2006 ca. 75.000 Euro aufgewendet werden müssen. Das tatsächliche Potenzial werden wir allerdings erst mit Durchführung einer technischen Optimierung im Rahmen eines allgemeinen Portal-Redesigns gegen Ende 2007 realisieren.

8 ENABLING-SUCHE

Das letzte Feld der Aktivitäten von T-Online befasst sich mit der Nutzung von Suchmaschinentechnologie im Kontext von Produkten und Prozessen außerhalb des Suchportals („Enabling“). Ursprünglich lag das Enabling nicht allzu stark im Fokus unserer Bemühungen, ging es uns zunächst doch vor allem darum, weitere Inhalte der Portalwelt im Kontext des Suchportals zugänglich zu machen, um so unser Produkt besser gegenüber dem Wettbewerb zu differenzieren. Im Sommer 2006 betrachteten wir aus dieser Motivation heraus erstmals die Option, das populäre Musikdownload-Angebot „Musicload“ als eigenen vertikalen Suchbereich in unser Portal aufzunehmen. Musicload verfügt über eine klassische Datenbank-Suche, die nicht ganz unseren Vorstellungen entsprach. Also indexierten wir einen XML-Dump der Produktdaten (einen Daten-Export im XML-Format) mithilfe unserer Enterprise Search und entwickelten zusätzliche Perspektiven auf den Inhalt, welche wir über sogenannte Cluster zugänglich machten. Mit einer vergleichsweise geringen Investition realisierten wir auf diese Weise eine Suche über den von Musicload angebotenen Produktkatalog, die der im Musicload-Portal integrierten Suche deutlich überlegen war, da sie zusätzliche Filter anbot, die den Musikliebhaber bei der Suche nach Titeln und Künstlern unterstützten. Erhielt man bei Musicload nach Eingabe des Begriffs „Davis“ noch hunderte von Ergebnissen in nicht nachvollziehbarer Reihenfolge, die zudem nur in Dezimalschritten angezeigt wurden, so offerierte die indexbasierte Suche dem Nutzer gleich Cluster wie „Pop“, „Jazz“, „Klassik“ oder „Miles Davis“, „Spencer Davis Group“, wobei die Cluster mit den meisten Einträgen oben standen und somit die Popularität bzw. auch die Schaffenskraft einzelner Künstler berücksichtigte. Schnell

wurde uns klar, dass die Index-basierte Suche der Datenbanksuche überlegen ist, da sie unabhängig vom Datenmodell auch Volltext-Inhalte erfassen und clustern kann. In weiteren Projekten wurden eine Auto-Suche (basierend auf einem XML-Dump von Autoscout24) sowie eine Suche für das Gamesload-Portal realisiert. Schnell reihte sich Projekt an Projekt, auch das Einsatzgebiet wechselte. Neben einer speziellen „Hilfe“-Suche, die ausschließlich Support-bezogene Inhalte unseres Portals erschloss, setzten wir auch einen Piloten auf, der Informationen unterschiedlicher Quellen (Datenbanken, Manuals, Webseiten) aus Sicht eines Callcenter-Mitarbeiters aufbereitet, wodurch ihm parallele Abfragen – und, dadurch bedingt, längere und kostenintensivere Bearbeitungszeiten je Call – erspart wurden.

Aktuell befinden sich auch Umsetzungen für Musicload (in Verbindung mit einem Data Cleansing der Metadaten) und Softwareload auf dem Weg. Innerhalb eines einzigen Jahres haben wir uns mit unserem „Enabling-Service“ quer durch die Portal-Landschaft von T-Online gearbeitet und den Mehrwert von Suchtechnologie demonstriert.

9 FAZIT

Die zentrale Erkenntnis aus diesen Umsetzungsbeispielen ist, dass die Online-Suche ohne jede Frage eine Schlüsseltechnologie ist. So wie wir heute bei T-Online ein großes Content-Management-System (CMS) betreiben, auf welchem ganz selbstverständlich alle unsere Portalangebote aufsetzen, so sehen wir inzwischen auch den Bedarf, eine zentrale Suche-Plattform vorzuhalten, die grundsätzlich für jedes Projekt verfügbar ist und jederzeit angesprochen werden kann, um die Realisierung von Online-Geschäftsmodellen, aber auch die Optimierung von internen Geschäftsprozessen zu unterstützen.

Die Online-Suche hat mit all ihren Facetten, ob nun in Form von Suchmaschine, SEO, SEM oder Enabling, über die letzten Jahre massiv an Bedeutung gewonnen und das Online-Geschäft um neue, erfolgreiche Geschäftsmodelle bereichert, für mehr Relevanz und Struktur gesorgt und einen jederzeitigen Zugriff auf nahezu jede frei zugängliche Information im Web ermöglicht. Online-Suche hat damit die Rahmenbedingungen für die erfolgreiche Umsetzung von Geschäftsideen im Internet nachhaltig und entscheidend geprägt. Dennoch darf man davon ausgehen, dass die Entwicklung noch lange nicht am Ende ist. Fest steht jetzt schon, dass in Zukunft nur erfolgreich sein kann, wer die Mechanismen, die von der Suchtechnologie ausgehen, versteht und ihre Anwendung beherrscht.

10 LITERATUR

AGOF 2006

AGOF e.V.: *internet facts 2006-IV*. URL: <http://www.agof.de/graphiken-if-2006-iv.download.91ca9bd86708b7ce3561c7e4ff753758.pdf> [Stand: 12.09.2007].

OVK 2007

Online-Vermarkterkreis: *OVK Online-Report 2007/01*. URL: http://www.bvdw.org/file-admin/downloads/marktzahlen/basispraesentationen/OVK_OnlineReport_200701.pdf [Stand: 12.09.2007].

Webhits 2007

„Web-Barometer“, in: www.webhits.de [Stand: 12.09.2007].

> VOLLTEXTSUCHDIENSTE – WAS DARF/SOLL WIE HINEIN UND HINAUS? EINIGE ANMERKUNGEN ZUM OPIUM DER INFORMATIONSGESELLSCHAFT¹

RAINER KUHLEN

1 ZUSAMMENFASSUNG

Auf den wissenschaftlichen Fachinformationsmärkten sind einerseits erhebliche *Konvergenz- und Integrationseffekte* in den Leistungen und Angeboten auszumachen, also eine Integration *klassischer* Verlags-, Bibliotheks- und Dokumentations- oder Fachinformationsleistungen sowie der neuen Angebote der post-professionellen Akteure auf den Internet-Märkten. Gemeint sind die Googles, Yahoos etc. sowie Dienste wie Amazon, aber auch Wikipedia. Auf der anderen Seite sind hoch kompetitive institutionelle *Desintegrationseffekte* bei den Angebotsstrategien festzustellen. Insbesondere zeichnen sich erhebliche Dismediations- und Retailtendenzen ab, die die Leistungen klassischer Mittel erodieren lassen. An ausgewählten Beispielen – AutoDoc FIZ Karlsruhe, Dokumentlieferdienst subito, Onleihe der öffentlichen Bibliotheken sowie das ScienceDirect- und Scirus-Angebot von Elsevier, erweitert um den verlagsübergreifenden CrossRef-Dienst – werden gegenwärtige Dokumentnachweis- und Lieferdienste kurz beschrieben und dann den Diensten der neuen Akteure, exemplarisch den beiden Varianten von Google Book Search, gegenübergestellt. Im Anschluss daran werden gegenwärtige Trends aufgezeigt, auch im Kontext der Regulierungen durch das Urheberrecht und dessen Modifikation durch Lizenzverträge in Public Private-Kooperationen. Trotz oder wegen des „Opiums der Googles“ – süß, gebührenfrei, wenn auch nicht kostenlos, verführerisch, süchtig und abhängig machend – wird die Informationsversorgung in Bildung und Wissenschaft weiter als primär öffentliche Aufgabe gesehen.

2 IN RICHTUNG DISMEDIATION UND RETAILGESCHÄFT

Mit „Volltextsuchsystemen“ im engeren Sinne sind hier elektronische Nachweissysteme gemeint, die Volltexte, bevorzugt Bücher, nachweisen und vollständig oder teilweise anzeigen sowie in der Regel auch Information darüber geben, wie man diese Volltexte entweder über Bibliotheken einsehen oder über die entsprechenden Verlage kaufen kann. Neben der knappen Darstellung der zentralen Funktionen und Leistungen dieser Systeme soll hier vor allem auf deren Auswirkungen auf das Wissenschaftssystem bzw. auf

¹ Dieser Text ist zum Abdruck im von Friedemann Mattern herausgegebenen acatech Tagungsband *Wie arbeiten die Suchmaschinen von morgen?* freigegeben. Ansonsten gilt die folgende Creative-Commons-Lizenz:



Dieses Dokument wird unter folgender [Creative-Commons-Lizenz](http://creativecommons.org/licenses/by-nc-sa/2.0/de/)

veröffentlicht: <http://creativecommons.org/licenses/by-nc-sa/2.0/de/>

2008 wird dieser Text außerdem in erweiterter Version als Kapitel einer geplanten Publikation des Verfassers mit dem Arbeitstitel „Erfolgreiches Scheitern – eine Götterdämmerung des Urheberrechts?“ erscheinen.

den allgemeinen Umgang mit Wissen und Information eingegangen werden. Behandelt werden ebenfalls einige urheberrechtliche Aspekte der Volltextsuchsysteme.

Im weiteren Sinne gehören zu solchen Systemen natürlich auch die *Volltextdatenbanken der Online-Datenbanken* des internationalen (vor allem) wissenschaftlichen Fachinformationsmarktes, die als Referenz- und Volltextsysteme in der Regel Zeitschriftenbeiträge enthalten. Ebenso könnten dazu die in ihrer Funktionalität und Abdeckung sich ständig erweiternden *Suchmaschinen* sowie die privaten bzw. kommerziellen oder öffentlichen *Dokumentnachweis- und Volltextliefersdienste* gerechnet werden.

Entsprechend sind auf den wissenschaftlichen Fachinformations- bzw. -publikationsmärkten erhebliche *Konvergenz- und Integrationseffekte* in den Leistungen und Angeboten auszumachen. Das heißt, es vollzieht sich eine starke Integration klassischer Verlags-, Bibliotheks- und Dokumentations- oder Fachinformationsleistungen sowie der neuen Angebote der post-professionellen Akteure auf den Internet-Märkten. Das schließt nicht aus, dass gleichzeitig hoch kompetitive institutionelle *Desintegrationseffekte* bei den Angebotsstrategien festzustellen sind.

Diese Effekte lassen, wie auf allen Internet-Teilmärkten, die Leistungen der klassischen Mittler erodieren. Auch wenn die Bibliotheken und Informationszentren weiter wichtige Mittlerdienste anbieten, ist unverkennbar, dass die professionellen kommerziellen und semi-kommerziellen Mittlerleistungen sich zunehmend in *dismediatisierte Retaildienste* wandeln. Das heißt, sie umgehen sozusagen die klassischen Mittler wie Bibliotheken, Information Broker, Buchhändler, Musik- und Video-Shops und wenden sich unter Vermeidung bzw. Reduktion bisheriger Transaktionskosten direkt an die Endkunden. Das trifft nicht nur für die Publikumsmärkte zu, sondern auch für Fachinformations- und -kommunikationsmärkte.

Mit den „postprofessionellen Akteuren“ aus der Kapitelüberschrift sind die Googles, Amazons, Wikipedias mit ihren ausdifferenzierten Angeboten gemeint, die längst über die Leistungen von Suchmaschinen, Buchbestelldiensten und Enzyklopädiën hinausgegangen sind. Diese haben ihre originale Kompetenz nicht wie die klassischen Wissenschaftsverlage und Bibliotheken bzw. Fachinformationszentren durch den engen Kontakt mit Bildung und Wissenschaft erworben. Dies hat Konsequenzen nicht zuletzt für Qualitätssicherung und Organisationsformen, aber auch für urheberrechtliche Besitzansprüche. Viele dieser Akteure haben ursprünglich Verhaltensformen, durchaus auch normative Einstellungen (s. unten), im Umfeld des Internets der Netizens (also aus prä-kommerzieller Internet-Zeit) entwickelt. Das schließt nicht aus, dass sie sich dann immer mehr mit ihren Angeboten an den allgemeinen Publikumsmärkten und, im Zuge der fortschreitenden Kommerzialisierung des Internets allgemein, an Gewinnerwartungen orientiert haben. Letzteres gilt nicht für alle, wie das Beispiel Wikipedia zeigt.

In methodischer Hinsicht hat sich in manchen Diensten der neuen Akteure auch ein neues Verständnis von „Werk“, „Autor“ und „Qualität“ ausgeprägt, das sich nicht mehr an Prinzipien wie Individualität, Subjektivität, Originalität (bezüglich der Produzenten) oder Autorität und auch nicht an Vorstellungen eines geschlossenen oder abgeschlossenen Produkts orientiert, sondern bei dem das in den Vordergrund gerückt wird, was mit Merkmalen wie „Kollaboration“, „Hypertextifizierung“ und „Virtualisierung“ beschrieben werden kann. Das wird besonders deutlich bei Wikipedia, aber auch bei den neuen sozialen Diensten im Web 2.0-Paradigma, bei denen sich auch Suchmaschinen wie Google und Yahoo stark engagieren. Es ist nicht zu übersehen, dass sich mit *neuen normativen Einstellungen* zu Autor, Werk und Qualität – um nur bei diesen drei grundlegenden Kategorien des wissenschaftlichen Publikationsgeschehens zu bleiben – auch neue Organisationsformen und neue Geschäftsmodelle sowie neue Einstellungen zur Reichweite und Gültigkeit von Regelungen des Umgangs mit Wissen und Information entwickeln.

Gegenwärtig sind wir allerdings wohl noch eher in einer Übergangsphase, in der unter Rückgriff auf die bislang erstellten und exklusiv durch Bibliotheken und Verlage bereitgestellten Werke neue Nachweis-, Anzeige- und Lieferdienste entstehen. Hierdurch werden in der Regel keine neuen Wissens- oder Informationsobjekte, sondern eher *Meta-dienste* erzeugt. Diesen Diensten – nicht zuletzt, weil sie in der Regel auf die rasche Erzeugung großer Quantitäten setzen (im Prinzip, wie Google, auf das gesamte Wissen der Welt) – sind die klassischen aufwändigen, bibliothekarisch-informationsmethodischen Erschließungs-, Aufbereitungs- und Qualitätssicherungsverfahren eher Belastung und Ballast, die mit Blick auf Endnutzerbedürfnisse vermeintlich vernachlässigt werden können.

Dass durch den Rückgriff auf bisherige Bestände Urheberrechtsprobleme entstehen, ist offensichtlich. Diese werden nicht dadurch entschärft, dass diese Dienste nicht direkt auf einen kommerziellen *Return of Investment* abzielen – die Nutzung der Dienste der Googles ist gebührenfrei –, sondern in erster Linie auf die kommerzielle Vermarktung der durch die Nutzung dieser Dienste entstehenden Daten der Nutzer (Interaktions-, Transaktions-, Nutzungsdaten).

Zu vermuten ist, dass der besonders von Google bei Google Book Search (GBS) gepflegte, quasi nonchalante, zumindest in den USA aber offenbar durch das *Fair-use-Prinzip* gedeckte Umgang mit Wissensobjekten, an denen andere Rechte haben (s. unten), einen Einfluss auch auf unsere Einstellung zu öffentlich gemachtem Wissen haben wird. Dieser Einstellungswandel ist ja schon vor einigen Jahren durch die Musikausbörsen Napster eingeleitet worden. „Napsterisierung“, konsequent zu Ende gedacht, ist der Vorgang der weitgehend ungehinderten Aneignung und Verteilung von Wissen bzw. des Teilens von einmal erworbenen Wissensobjekten. Napsterisierung könnte heute durch „Googleisierung“ und „Wikipediatisierung“ zum allgemeinen Prinzip der Produktion, des Nachweises und der Nutzung von Wissensobjekten jeder Art werden, und das nicht nur auf den Publikumsmärkten, sondern auch auf den Märkten der Wissenschaftskommunikation.

Information in elektronischen Umgebungen will frei sein. Dass es dagegen massive Widerstände der kommerziellen Copyright-Industrien gibt, unterstützt von den Urheberrechts- bzw. Copyright-Regelungen der letzten Jahre, ist ebenso klar. Ebenso erkennbar ist aber auch, dass durch die „freien“ Dienste, zumal, wenn sie von kommerziell ausgerichteten Organisationen übernommen und an Verwertungsbedürfnisse angepasst worden sind, kein Paradies der „Wissengesellschaft“ entstehen wird. Kein Unternehmen verschenkt etwas, wenn es nicht, über welche Umwege auch immer, von den „freien“ Geschenken letztlich doch Gewinne erwartet. Google macht das im großen Stil vor, und es sieht ganz so aus, dass das Bezahlen mit anderer Währung als mit Geld – also mit persönlichen Daten oder dem Offenlegen von Interessen und Verhaltensformen – möglicherweise sogar zu größeren Verlusten führt.

Ich habe von einer „Übergangsphase“ gesprochen. Auf die weitere Entwicklung auf den Publikationsmärkten, die ich mit „Kollaboration“, „Hypertextifizierung“ und „Virtualisierung“ angesprochen habe und die durch die Konzepte des „individuellen Autors“, des „geschlossenen Werkes“ oder der durch Autorität zertifizierten „Qualität“ infrage gestellt werden könnten, kann ich hier nicht näher eingehen. Bleiben wir also bei den anfangs kurz definierten Volltextsuchsystemen, bei denen ich mich vor allem auf GBS einlassen möchte (ohne die Bedeutung anderer Volltextsuchdienste wie die der Open Content Alliance oder von Amazon schmälern zu wollen). Vorab möchte ich jedoch kurz und nur exemplarisch auf einige Nachweis- und Liefersysteme der professionellen Akteure (Verlage, Bibliotheken, Fachinformationszentren) eingehen, die ebenfalls zunehmend auf Endnutzer abzielen und ebenfalls, direkt oder indirekt, Volltexte nachweisen und ausliefern (lassen).

3 IM KONTEXT ANDERER VOLLTEXTNACHWEIS- ODER LIEFERDIENSTE

Natürlich ist hier keine annähernd vollständige Informationsmarktanalyse intendiert, weder mit Blick auf die kommerziellen Anbieter und Angebote, noch auf die immer mehr den Potenzialen elektronischer Technologien Rechnung tragenden Bibliotheksleistungen. Exemplarisch stelle ich nur einen hier einschlägigen Dienst eines Fachinformationszentrums (FIZ AutoDoc) vor (3.1), eine Dokumentlieferleistung aus dem Bibliotheksbereich (subito, in 3.2), einen derzeit kontrovers diskutierten elektronischen Ausleihdienst einer öffentlichen Bibliothek (Onleihe, in 3.3) und als Beispiel für eine komplexe Verlagsdienstleistung ScienceDirect und Scirus von Elsevier (3.4) sowie die verlagsübergreifende Ausweitung durch CrossRef (3.5). Dem schließt sich die Darstellung des GBS-Dienstes des neuen Akteurs Google an. Zum Abschluss sollen einige Trends herausgearbeitet und Konsequenzen daraus gezogen werden.

3.1 AUTODOC – MEHRWERTEFFEKTE MIT HOHEN KOSTEN

Über FIZ AutoDoc vom Fachinformationszentrum Karlsruhe (oft auch als „Wissenschafts-host“ bezeichnet) können gegenwärtig ca. 150.000 Zeitschriftentitel nachgewiesen und ausgeliefert werden.² Der Dienst kann direkt aus der Online-Recherche mit den FIZ-Datenbanken aktiviert werden, ist aber auch als Stand alone-Angebot möglich. Die Mehrwerteffekte beziehen sich in erster Linie auf die Geschwindigkeit der Auslieferung: Die Standardauslieferungszeit beträgt bei FIZ AutoDoc 48 Stunden; im Rush-Dienst geht es innerhalb 24 Stunden; bei einer „Super Rush Order“ kann es auch innerhalb 3 Stunden gehen. Die Dokumente können elektronisch als PDF, mit DRM-FileOpen kontrolliert, via E-Mail ausgeliefert, aber auch als Fax oder per Post geschickt oder vom Anbieter heruntergeladen werden. Den Mehrwerten stehen weitaus höhere Gebühren als beim klassischen Leihverkehr der Bibliotheken (oder früher bei subito) gegenüber. Fachinformationszentren müssen nach den Vorgaben ihrer Geldgeber (in diesem Fall überwiegend das Bundesministerium für Bildung und Forschung) einen möglichst hohen Kostendeckungsgrad erreichen. Die Kunden von AutoDoc kommen sicher auch aus den Bereichen Bildung und Wissenschaft, aber wohl auch und zunehmend aus der Wirtschaft. Bei diesen Diensten werden die Urheberrechtsvorschriften selbstverständlich eingehalten, also auch entsprechende Abgaben an die Verwertungsgesellschaften bzw. individualisierte oder pauschale Lizenzgebühren an die Rechteinhaber der Volltexte gezahlt. Die Standard-PDF-Auslieferung kostete bei FIZ AutoDoc im Oktober 2007 9,95 Euro zuzüglich 5,58 Euro Copyright-Gebühr in Deutschland (bei europäischen Lieferanten nur 1,60 Euro), zuzüglich Mehrwertsteuer; beim Super Rush Order-Dienst liegen die Kosten bereits bei 47,20 zuzüglich der genannten Gebühren, also insgesamt bei ca. 60 Euro.

Ansonsten haben Fachinformationszentren Verträge mit Depot- oder Zentralbibliotheken abgeschlossen – in Deutschland sind das für den Fachkommunikationsmarkt in erster Linie die Technische Informationsbibliothek Hannover oder die Zentralbibliothek Medizin in Köln –, sodass Benutzer direkt aus der Online-Recherche mit einem speziellen Kommando die für sie relevanten Dokumente bestellen können. Nach der im Juli 2007 beschlossenen zweiten Urheberrechtsnovellierung wird die Lieferung elektronischer Dokumente (und sei es nur in grafischer Form) durch die angeforderten Bibliotheken schwieriger oder gar unmöglich werden, da der Gesetzgeber über § 53a UrhG der Verlagswirtschaft quasi ein Monopol für den elektronischen Dokumentenversand zugesteht, wenn bestimmte Bedingungen erfüllt sind. Zu diesen Bedingungen gehört auch, dass es „offensichtlich“ sein muss, dass für ein nachgefragtes Dokument ein kommerzielles Angebot auf dem Markt existiert. Dies wird im wachsenden Retailgeschäft immer mehr der Fall sein (vgl. Abschnitte 3.4 und 3.5). Ob dies Kunden aus der Wirtschaft beeinträchtigen wird, muss sich noch zeigen. Für Bildung und Wissenschaft werden solche Dienste damit vermutlich uninteressant bzw. unerschwinglich.

² <http://autodoc.fiz-karlsruhe.de/>

3.2 DOKUMENTLIEFERUNG IM BIBLIOTHEKSVERBUND

subito, aus einer Bund-Länder-Initiative hervorgegangen, versteht sich als Dachorganisation, die die Belange der Mitgliedbibliotheken hinsichtlich der angebotenen elektronischen Informationsdienste vertritt und diese Dienste bereitstellt. subito bietet eine gezielte Titelrecherche zu wissenschaftlichen Dokumenten in seinen Katalogen an und erlaubt bei einschlägigen Treffern die Bestellung einer Aufsatzkopie aus einer Zeitschrift oder auch die Bestellung eines Buches zur Ausleihe. Nach der Juli 2007 im Bundestag beschlossenen zweiten Urheberrechtsnovellierung hat subito seine Produkt- und Preispolitik für drei Bezugsgruppen ändern müssen.

- a) Für den deutschen subito Direct Customer Service hat der subito-Vorstand Juni 2007 beschlossen, sich zunächst ganz aus dem elektronischen Lieferdienst (für Deutschland) zurückzuziehen. Post- und Fax-Versand sind jetzt auch nach den Regelungen des Zweiten Korbs 2007 in § 53a UrhG grundsätzlich erlaubt. Dafür hat subito die Preise gesenkt. Nicht-kommerzielle Nutzer im Normaldienst zahlen pro Artikel um 5 Euro für E-Mail-Versand, 8 Euro für Post und Fax. Für den Eildienst variieren die Preise je nach Lieferbibliothek. Sie sind im Durchschnitt jeweils um 3 bis 5 Euro höher als beim Normaldienst. Bei kommerziellen Kunden berechnen die Bibliotheken zwischen 10 und 15 Euro, privaten Kunden etwas weniger.³
- b) Auch beim internationalen subito Direct Customer Service wird zwischen nicht-kommerziellen und kommerziellen Nutzern unterschieden. Der Service ist registrierungspflichtig. Für die lizenzpflichtigen Dokumente hat subito im Juli 2006 einen entsprechenden Vertrag mit STM abgeschlossen.⁴ Voraussichtlich müssen je nach Verlag zwischen 24 und 33 Euro pro Artikel von den Kunden gezahlt werden, also auch von den im nicht-kommerziellen Bereich angesprochenen Auszubildenden, zum Beispiel bei der *Acta Informatica* für alle drei Lieferformen 24,50 Euro bzw. bei der *Acta Anaesthesiologica Scandinavica* 33 Euro. Diese Lizenzverträge für die Dokumentlieferdienste von subito finden in den Ländern Deutschland, Österreich, Schweiz und Liechtenstein (bislang) keine Anwendung (vgl. Anm. 3). Dokumente (als grafische Datei oder PDF) aus den Lizenzverträgen sind über ein DRM-System verschlüsselt und so geschützt. Es gelten unter anderem die folgenden Regelungen:
 „Der Kunde darf das Dokument zehnmal auf ein- und demselben Rechner ansehen und einmal ausdrucken. [...] Nach Ablauf eines Monats nach dem Versand der E-Mail kann das Dokument nicht mehr angesehen und nicht mehr

³ subito verhandelt derzeit mit Blick auf die Versorgung von Kunden aus den Hochschulen und den außer-universitären Forschungsbereichen mit „dem Börsenverein und der Vertretung der großen technisch-naturwissenschaftlichen Verlage STM Stichting. Die PDF-Lieferung gescannter Dokumente innerhalb Deutschland, Österreich, Schweiz und Liechtenstein wird dann gegen Bezahlung einer Lizenzgebühr auch in den deutschsprachigen Ländern wieder möglich werden“. (Mitteilung aus Sparte Urheberrecht bei www.subito-doc.de).

⁴ STM ist die International Association of Scientific, Technical and Medical Publishers. S. <http://www.stm-assoc.org/>.

gedruckt werden. [...] Es ist nicht gestattet, das Wasserzeichen zu entfernen. Dem Bibliotheksbenutzer ist zusammen mit dem Dokument der Urheberrechtshinweis auszuhändigen. [Zur] Ansicht und zum Druck der geschützten Dokumente ist die einmalige Installation eines Plugin notwendig. [...] Bitte beachten Sie, dass das Plugin nur bei einer offenen Internet-Verbindung arbeitet. Sie können DRM-geschützte Dokumente also nur bei gleichzeitig aktivem Internet-Zugang öffnen und ausdrucken.“

- c) Schließlich wendet sich subito Library Service auch an die Bibliotheken, und zwar an solche, die zu mindestens 52 Prozent aus öffentlichen Mitteln finanziert werden, im gesamten, auch deutschsprachigen internationalen Bereich. Ein Eildienst wird hier nicht angeboten. Der Dienst gilt nicht für kommerzielle Kunden. Bei der Preispolitik und dem Service wird allerdings zwischen dem deutschsprachigen und sonstigen internationalen Bereich unterschieden; derzeit werden im ersteren Bereich 4 Euro für den E-Mail-Grundpreis sowie 8 Euro für Fax und Post fällig, im zweiten jeweils geringfügig mehr. Es kann eine Lizenzgebühr von 3,50 Euro dazukommen. Diese Gebühr wird fällig, wenn die Anfrage einen Artikel aus einem Verlag betrifft, mit dem subito einen Lizenzvertrag abgeschlossen hat (vgl. die Regelungen unter b)).

An dem subito-Dienst ist bemerkenswert, dass eine zunächst mit öffentlichen Mitteln eingerichtete Dienstleistung, die durchaus große Benutzerakzeptanz gefunden hat, durch die Urheberrechtsgesetzgebung gezwungen wurde, nicht mehr allein auf die Kooperation mit den öffentlichen Bibliotheken zu setzen, sondern mit dem Börsenverein und STM Lizenzverträge abzuschließen. Damit beugt sich auch subito zwangsläufig den Auflagen des Urheberrechts, das ja über § 53a der Informationswirtschaft quasi ein Monopol für die Dokumentlieferung gewährt. Nur über Lizenzvereinbarungen ist subito berechtigt, Zeitschriftenartikel zu vervielfältigen und sie, DRM-geschützt, elektronisch zu versenden. Nur auf diesem Wege kann subito weiterhin seinem Auftrag gerecht werden, internationale wissenschaftliche Publikationen aus Forschung und Lehre bereitzustellen.⁵

3.3 NEUE DIENSTLEISTUNGEN AUCH IN ÖFFENTLICHEN BIBLIOTHEKEN – DAS KONZEPT „ONLEIHE“

Bibliotheken experimentieren natürlich mit Formen der Dokumentbereitstellung, die über die klassischen Formen wie Fernleihe weit hinausgehen. Exemplarisch sei hier auf das Projekt Onleihe eingegangen.⁶ Die Idee, Online-Materialien auszuleihen, ist zunächst verblüffend, denn die Konzepte „online“ und „Ausleihe“ scheinen nicht recht zusammenzupassen. Einige öffentliche (städtische) Bibliotheken haben dies in Zusammenarbeit mit der Firma DiViBib jedoch zu realisieren unternommen. DiViBib ist ein Produkt der

⁵ Mehr Informationen dazu finden sich bei subito 2006.

⁶ Vgl. den Telepolis-Artikel dazu bei Heise 2007.

ekz.bibliotheksservice GmbH in Reutlingen, ein traditioneller Dienstleister speziell für Bibliotheken.⁷ DiViBib unterstützt Bibliotheken zum einen beim Aufbau von digitalen Bibliotheksplattformen und zum anderen durch den Vertrieb von digitalen Medien, die Bibliotheken in ihr Angebot aufnehmen können.

Im letzteren Kontext ist das Produkt Onleihe entstanden. DiViBib vermarktet die Produkte nicht selber, sondern bietet den Service öffentlichen Bibliotheken an. Auf der Website von DiViBib wird auf einer Deutschlandkarte angezeigt,⁸ welche Stadtbibliotheken schon den Ende Mai 2007 (mit Hamburg und Würzburg) gestarteten Onleihe-Service bereitstellen. Nutzer des Dienstes müssen bei der jeweiligen Bibliothek registriert sein.⁹ Die Nutzung selbst erfolgt selbstverständlich standortunabhängig über das Internet. Bislang ist die Nutzung für die bei der Anbieterbibliothek registrierten Nutzer nicht mit weiteren Gebühren verbunden.

Als elektronische Medien werden E-Books, Audio-Files (von Texten, Vorlesebüchern oder Musik) und Video-Files angeboten. Die Bibliotheken wählen aus dem Angebot, für das DiViBib mit den Verlagen die entsprechenden Nutzungsrechte erworben hat, die Objekte aus, die sie in ihrer lokalen virtuellen Stadtbibliothek haben wollen. Zur Startphase Ende Mai 2007 wurden von den ersten Bibliotheken (wie Würzburg und Hamburg) etwa 10.000 Objekte zur Verfügung gestellt.

Entsprechend dem Leihprinzip werden die Objekte für eine fest vereinbarte Zeit zur Nutzung bereitgestellt. Ist die Zeit abgelaufen, kann das Objekt nicht mehr genutzt werden und ist damit – in der klassischen Ausleihterminologie – „zurückgegeben“. Da die Rückgabe automatisch mit Fristenende geschieht, gibt es auch keine Mahngebühren mehr. Nach den Vereinbarungen mit den Rechteinhabern an den auszuleihenden Objekten ist bislang vorgesehen, dass das jeweilige E-Objekt nur in der Anzahl gleichzeitig ausgeliehen werden darf, die der Anzahl der Rechte an dem Objekt bzw. der Menge der Exemplare entspricht, die die Bibliothek erworben hat. Ist das E-Objekt ausgeliehen, kann man es sich vormerken lassen und bekommt auch angezeigt, wie viele Vormerkungen schon vorhanden sind.

Diese restriktive Regelung für eine serielle Nutzung besteht jetzt auch für die Bereitstellung von digitalen Objekten an den Leseplätzen der Bibliotheken nach § 52b UrhG.¹⁰ Eine solche Regelung macht, unabhängig von den durch das Urheberrecht geschützten Verwertungsinteressen, im elektronischen Umfeld sicher keinen Sinn und kann nur als lästiges, nutzerunfreundliches Verknappungs- und, hoffentlich, nur als Übergangsphänomen verstanden werden.

Andere Leistungen sind durchaus als Mehrwertleistungen einzuschätzen. So werden Preview-Leistungen für alle medialen Objekte angeboten, das heißt man kann sich von der Video-Datei eine kurze Sequenz anschauen, in ein Musikstück hineinhören oder

⁷ Neben ekz.bibliotheksservice GmbH ist Holger Behrens Gesellschafter von DiViBib. Vgl. DiViBib 2007.

⁸ Vgl. *bibliothek digital* 2007.

⁹ Vgl. die Nutzung bei dem ersten Anbieter von Onleihe, der Stadtbibliothek Würzburg. Stadt Würzburg 2007.

¹⁰ Die Anzeige gemäß dem Prinzip der Bestandsakzessorietät darf nach den Erläuterungen zu § 52b UrhG nur bei besonderen Belastungsspitzen, wie zum Beispiel bei anstehenden Prüfungen, durchbrochen werden, bei denen auch ein vorhandenes Exemplar an bis zu vier Leseplätzen elektronisch eingesehen werden darf.

durchaus in einigen Seiten aus einem E-Book stöbern, bevor man sich zu der Onleihe entscheidet. Auch kann man sich im Ausgang von einem selektierten Titel anzeigen lassen, welche anderen Objekte der jeweilige Verlag anbietet oder was der betreffende Autor, Interpret etc. sonst noch veröffentlicht hat.

Die Verwaltung der gesamten Vorgänge übernimmt das Digital Rights Management (DRM). Hier beginnt das Ganze problematisch zu werden, auch wenn die Anwender von Onleihe DRM quasi als naturgegeben bzw. als unvermeidbar bei Anwendungen ansehen, bei denen elektronische Objekte über Lizenzvereinbarungen Nutzern bereitgestellt werden sollen.¹¹ DRM regelt im einzelnen, welcher Nutzungsspielraum gegeben wird, zum Beispiel ob der Text neben der Anzeige auf dem eigenen Bildschirm auch ausgedruckt bzw. auf einen anderen elektronischen Speicher abgelegt oder gebrannt werden kann. Die Nutzungsbedingungen im Detail werden in der Regel über die Verträge der bereitstellenden Bibliotheken mit den Rechteinhabern, also den Verlagen festgelegt und sind für die Nutzer nach der Selektion eines bestimmten Titels sichtbar. Die Benutzernummer wird in ein digitales Wasserzeichen eingearbeitet. Bei Audio- und Video-Files ist im Allgemeinen vorgesehen, dass sie nicht nur auf dem eigenen Computer, sondern auch auf DRM-fähigen MP3-Playern abgespielt werden können. Zudem sind wegen der Zusammenarbeit mit Microsoft (das DRM wurde von Microsoft kostenlos zur Verfügung gestellt) alle Objekte nur im proprietären Windows Media Format (WMF) nutzbar, sodass nicht nur die Apple oder GNU/Linux-Nutzer ausgeschlossen sind, sondern – und das hat mit Blick auf die jugendlichen Nutzer in öffentlichen Bibliotheken durchaus auch quantitative Relevanz – auch die Besitzer von nicht DRM-/WMF-kompatiblen MP3-iPods. Allerdings ist auch diese Beschränkung vermutlich eher als Übergangsproblem zu sehen, da ja auch in anderen Bereichen sehr bald Software entwickelt werden wird, die derartige Beschränkungen außer Kraft setzt.

Dennoch: Dass sich Bibliotheken und Dienste wie subito, nicht zuletzt unter dem Druck der Urheberrechtsregulierungen, zunehmend auf DRM einlassen, ist problematisch genug. Dass sie zu diesem Zweck Vereinbarungen mit Anbietern proprietärer Software eingehen, durch die diesen Anbietern Monopolrechte zugebilligt werden, ist aufgrund der Finanzsituation der Bibliotheken nachzuvollziehen, aber dennoch nicht akzeptabel, zudem dadurch ja auch Bindungen für den Erwerb von Hard- und Software in der Zukunft entstehen. Bibliotheksdienstleistungen sollten prinzipiell offen sein und daher durchgängig nicht-prorietäre Software verwenden.¹²

¹¹ „Ein digitaler Ausleihvorgang, also das temporäre Zur-Verfügung-Stellen von Inhalten als Alternative zur dauerhaften Verfügbarkeit, [könne] nur über ein DRM abgebildet werden [...]. DRM [sei] die Grundvoraussetzung dafür [...], das von öffentlichen Bibliotheken praktizierte Ausleihen von Medien in die digitale Welt zu übertragen“, so Christian Hasiewicz, Bibliothekarischer Direktor bei DiViBib in einem Beitrag in INETBIB vom 10.9.2007.

¹² Angesichts dieses Erfordernisses fällt die Konsequenz besonders negativ auf, dass es für die Bibliotheken wie auch für öffentliche Informations- und Dokumentationsdienstleister keine Infrastruktur mehr gibt, durch die solche Leistungen hätten erbracht werden können. Die Gesellschaft für Information und Dokumentation (GID) wurde Ende der 80er Jahre aufgelöst, ein Jahrzehnt später das zentrale Bibliotheksinstitut. Und Ende 2006 wurde das bund-/länderfinanzierte Fraunhofer-Institut für Integrierte Publikations- und Informationssysteme (IPSI) geschlossen, das die Forschungsaktivitäten der GID erfolgreich fortgesetzt hatte. Fachinformationsinfrastruktur hat keine politische Unterstützung in Deutschland mehr. Das über dreißig Jahre lang aktive Referat im Forschungsministerium, das für Fachinformation zuständig war, wurde ebenfalls Ende 2006 aufgelöst.

Das Geschäftsmodell von DiViBib bzw. die Idee von Onleihe ist aber im Prinzip unabhängig von der konkret zur Anwendung kommenden Software und dürfte zu weiteren Experimenten mit Onleihe-Diensten führen. Weitere derartige elektronische Dienstleistungen sollten sich aber nicht auf die Umarmungs- oder Geschenkstrategien proprietärer Anbieter einlassen; und wenn sie es doch meinen tun zu müssen, sollten nach einer Übergangsphase Ausstiegsmöglichkeiten bzw. der Einstieg in offene Lösungen vereinbart und dann auch wahrgenommen werden.

3.4 RETAILDIENST VON SCIENCEDIRECT UND SCIRUS

Elsevier hat den beschriebenen Einstieg in das Retailgeschäft längst vollzogen. Schon bis 2004 hatte Elsevier 200 Millionen britische Pfund in ScienceDirect und andere Programme zur retrospektiven Digitalisierung und Archivierung von Zeitschriften investiert. Das ScienceDirect-Angebot macht etwa ein Viertel der gesamten weltweiten wissenschaftlichen, technischen und medizinischen Information aus.¹³ Schwerpunkt sind dabei die über 2000 Elsevier-eigenen Zeitschriften, aber zunehmend auch Bücher und Buchserien. Neben den eigenen Zeitschriften zeigt ScienceDirect auch eine Vielzahl von Zeitschriften aus Beständen Dritter an. Daneben erschließt ScienceDirect über das Backfile-Programm auch alte Bestände – etwa 7,75 Millionen Artikel, davon ca. 4 Millionen von vor 1995, also Artikel, die elektronisch verfügbar gemacht wurden.

Die Suche nach Artikeln (oft auch über die Abstracts) ist kostenlos. Für die Bestellung eines Artikels muss, falls die Suche nicht von einem Vertragspartner wie einer Bibliothek gestartet wird, ein Account eingerichtet sein. Die Abrechnung erfolgt über Kreditkarte. Einige Zeitschriften sind gebührenfrei in der Nutzung, viele bieten freien Zugriff auf die älteren Bestände (meistens älter als 12-24 Monate).

ScienceDirect bietet laufend neue Mehrwertleistungen an. So können individuelle RSS-Feeds eingerichtet werden, damit man benachrichtigt werden kann, zum Beispiel wenn neue Artikel in speziellen Zeitschriften erschienen sind oder wenn ein bestimmter Artikel in einer Zeitschrift zu einem festzulegenden Zeitpunkt neu zitiert wird.

Elsevier erweitert seine Such- und Nachweisleistungen durch die auf Wissenschaft ausgerichtete Suchmaschine Scirus.¹⁴ Scirus hat über 450 Millionen Seiten (auch PDF) mit wissenschaftlichen Inhalten im Web indexiert, wobei nicht-wissenschaftliche Websites, die aber auch den jeweiligen Suchbegriff enthalten, ausgefiltert werden. Die meisten Seiten stammen von Websites mit dem „edu“-Suffix, gefolgt von „org“- und „com“-Sites. Für die Ergebnisanzeige verwendet Scirus ein Ranking-Verfahren, das sich zum Teil auf Vorkommenshäufigkeiten des gesuchten Begriffs in den Seiten, zum Teil auf dynamisches Ranking stützt, vergleichbar dem PageRank von Google (Ranking, in Abhängigkeit der Anzahl der Referenzierung von anderen Websites). Neben den offenen Websites indexiert Scirus eine Vielzahl wissenschaftlicher Ressourcen (Pre-Print-Server, Reports, Repositories, kommerzielle und öffentliche Nachweis- und Lieferdienste)¹⁵.

¹³ So die Eigenaussage von Elsevier, s. <http://www.sciencedirect.com/>.

¹⁴ www.scirus.com.

¹⁵ Zum Beispiel die E-Prints von ArXiv, die Dokumente von Digital Archives, PubMedCentral, ScienceDirect, aber auch so spezielle Ressourcen wie 12.000 Dokumente von CalTech oder 4.600 Volltexte der Humboldt-Universität Berlin. Eine vollständige Nachweisliste ist zu finden unter Scirus 2007.

Elsevier hat schon früh erkannt, dass Open Access eine für Verlage höchst bedrohliche und auch wohl unumkehrbare Entwicklung im wissenschaftlichen Publikationsbereich ist.¹⁶ Elsevier hat im Februar 2004 einen Kommentar zu „evolutions in scientific, technical and medical publishing and reflections on possible implications of Open Access journals for the UK“ vorgelegt,¹⁷ mit dem sich Elsevier, sicherlich über England hinaus anwendbar, auf die durch Open Access veränderte Marktsituation im wissenschaftlichen Kommunikationsbereich zu positionieren versucht hat. Darin werden alle Argumente vorgebracht, die bis heute die Einstellung der Verlagswelt zu Open Access prägen (vgl. die Brüsseler Erklärung von Anfang 2007 und die später gestartete PRISM-Initiative):

Der kommerzielle wissenschaftliche Markt funktioniert; er sei höchst leistungsfähig und begünstige die wissenschaftliche Innovationsrate; der Zugriff zum weltweiten Wissen sei für jeden Wissenschaftler gesichert; die Kosten für den Zugriff auf einen einzelnen Artikel seien kontinuierlich gesunken. Dagegen würde das Open Access-Modell mit einem „Author Pay“-Modell das Vertrauen der Öffentlichkeit in die Integrität und Qualität der wissenschaftlichen Publikationen unterminieren. Das Subskriptionsmodell, bei dem die Nutzer zahlen, sichere hohe Qualität und unabhängiges Peer Review hindere kommerzielles Interesse daran, Einfluss auf Veröffentlichungen zu nehmen (!). Open Access habe bislang kein verlässliches Finanzmodell entwickelt, die bislang verlangten Autorengebühren (maximal 1.500 britische Pfund) würden höchstens 40-60 Prozent der tatsächlichen Kosten decken, sodass die Öffentlichkeit, die Stiftungen oder Hochschulen hier einspringen müssten: Open Access-Herausgeber oder -Verleger würden nicht wie Elsevier in der Lage sein, in die Bereitstellung neuer Mehrwertleistungen zu investieren. Im Übrigen gehe es nicht an, dass die Regierungen durch Unterstützung von Open Access-Modellen verzerrend in das Marktgeschehen eingreifen.

¹⁶ Konstruktiv aufgegriffen hat unter anderem der Springer-Verlag die Open Access-Herausforderung und mit Open Choice einen eigenen Open Access-Dienst gestartet. Durch die Beteiligung am Springer Open Choice-Angebot (vgl. Springer 2007a) können Autoren entscheiden, ob sie ihre Beiträge Open Access stellen wollen oder sie dem Verlag zur kommerziellen Verwertung überlassen wollen. Falls sie sich für die erstere Option entscheiden, werden die Urheberrechte nicht exklusiv an den Verlag übertragen. Das Recht zur Veröffentlichung in der jeweiligen Springer-Zeitschrift wird natürlich gegeben – ob weitere Rechte für noch unbekannte Nutzungsarten gegeben werden, muss dann weiter auszuhandelnder Vertragsgegenstand sein. Bislang steht in der Autorenvereinbarung, dass die Autoren Springer das „exklusive Recht“ „to any commercial exploitation of the article“ übertragen. Die Autoren dürfen aber ihre Arbeiten zeitlich unbegrenzt selber archivieren und für jedermann öffentlich frei (gebührenfrei) zugänglich machen. Sie dürfen sie auch weiter verbreiten, das heißt, sie auch Open Access-Repositories überlassen, sofern es sich um eine nicht-kommerzielle Nutzung handelt und auf die originale Springer-Veröffentlichung referenziert wird. Mit Springer wird dazu eine Open Choice Licence vereinbart, die mit der Creative Commons-Lizenz kompatibel ist, („you are free to copy, distribute, display, and perform the work, to make derivative works“, vgl. Springer 2007b). Es handelt sich bei der Open Access-Option also nicht um eine Selbst- oder Direktpublikation, sondern um eine normale, Peer Review-kontrollierte und mit den gleichen von Springer bereitgestellten Mehrwerteffekten ausgestattete Veröffentlichung wie bei allen anderen Springer-Publikationen. Der Unterschied ist lediglich, dass die Autoren bzw. ihre Organisationen bzw. die Förderorganisationen, die die der Veröffentlichung zugrunde liegende Arbeit finanziert haben, für die Publikation des Artikels zahlen. Springer berechnet dafür derzeit 3.000 britische Pfund. Der Artikel wird über den Springer-Online-Dienst, SpringerLink (<http://www.springerlink.de>), für jeden Internet-Anschlussinhaber in der Welt zur Nutzung freigestellt. Für die Bibliotheken, die Springer-Zeitschriften subskribiert haben, reduziert sich der Subskriptionspreis für das folgende Jahr um den Anteil an Open Access-Publikationen in den jeweiligen Zeitschriften im aktuellen Jahr.

¹⁷ Elsevier 2007.

Wenig später im Jahr 2004 hat Elsevier dann ein erstaunliche Wende in seiner Geschäftspolitik vollzogen, die in gewisser Hinsicht durchaus als Anerkennung der Unumkehrbarkeit von Open Access angesehen werden kann, ohne dass damit aber der Anspruch auf die kommerzielle Verwertung zurückgenommen wird: Elsevier erlaubt seitdem all seinen Autoren, ohne Zeitverzögerung, also ohne Einfordern einer Embargo-Frist, ihre Artikel in der letzten, bei Elsevier eingereichten Version auf der eigenen Website oder der Website ihrer Institution einzustellen – mit Referenz auf die Website der Zeitschriftenveröffentlichung oder mit dem Vermerk der DOI (vgl. 3.5). Allerdings dürfe die öffentliche Bereitstellung nicht in der PDF- oder HTML-Version erfolgen, die aus ScienceDirect heruntergeladen werden kann, sondern nur in der Autoren-, zum Beispiel Word- oder Text-Textversion. Die Weitergabe des Artikels an ein Open Access Repository Dritter werde nicht gestattet.

Diese Liberalisierung wurde auch in der Open Access-Fachwelt durchaus ambivalent eingeschätzt. Während die einen das Festhalten am kommerziellen Modell grundsätzlich und die Einschränkungen speziell kritisierten, haben Open Access-Vertreter wie Peter Suber und Stevan Harnad die Öffnung von Elsevier positiv eingeschätzt – erstere aus der Überlegung, dass ein weiteres Entgegenkommen der Verlagswelt bei Fortbestand des kommerziellen Publikationsmodells nicht zu erwarten sei, die letzteren aus grundsätzlichen Erwägungen, dass nämlich der Green Road-Ansatz, also die Einstellung der Arbeiten durch die Autoren selber, durchaus ein vollwertiges und zukunftssträchtiges Open Access-Modell sei.

In jedem Fall ist mit diesen Auseinandersetzungen durch Elsevier eine Debatte über Geschäfts- und Organisationsmodelle in die Öffentlichkeit getragen worden, die unter Anerkennung der Open Access-Forderung nach gebührenfreiem Zugriff auf Wissen und Information dennoch deren kommerzielle Verwertung weiterhin möglich machen sollte.

3.5 CROSSREF

Die kommerziellen Elsevier-Angebote sind integriert in das verlagsübergreifende Cross-Ref-System, auf das sich seit dem Jahr 2000 so gut wie alle größeren wissenschaftliche Verlage verständigt haben¹⁸. Die Leistung von CrossRef beruht auf dem konsequenten Einsatz der DOI-Technik (Digital Object Identifier). Mit einem DOI kann jedes Informationsobjekt in elektronischen Umgebungen eindeutig referenziert werden und über die

¹⁸ Im Board of Directors zum Beispiel sind vertreten AAAS (Science), AIP, ACM, APA, Blackwell Publishers, Elsevier Science, IEEE, Wolters Kluwer, Nature, Sage, Springer, Taylor & Francis, Thieme, University of California Press, University of Chicago Press and Wiley.

zugeordneten Metadaten auch jede individuelle Information in einem Artikel, zum Beispiel – und das ist für CrossRef zentral – eine Literaturreferenz in einem wissenschaftlichen Artikel¹⁹.

Ein DOI bleibt fest mit dem spezifizierten Objekt verbunden, unabhängig von dem Ort, wo das Objekt gespeichert ist, und auch unabhängig vom aktuellen Besitzer. Da gerade auf den Informationsmärkten sehr häufig Zusammenlegungen und Aufkäufe von Informationsanbietern entstehen, besitzen im Laufe der Zeit Anbieter häufig DOIs, die wegen des konstant bleibenden DOI-Namens im Präfix-Teil auf Informationsanbieter verweisen, die gar nicht mehr als selbstständige Anbieter existieren. Die neuen Besitzer können allerdings neue Metadaten vergeben. Alle bei CrossRef vergebenen DOIs sind in der zentralen CrossRef-Datenbank registriert und können dort online eingesehen werden. CrossRef ist somit offizielle DOI-Link Registrierungsagentur. Die DOIs können auch in den gedruckten Versionen zum Beispiel von Zeitschriften eingesehen werden. Auch die Metadaten werden über CrossRef verwaltet und können daher von Suchmaschinen zur Indizierung eingesehen und verwendet werden. Mitte September 2007 waren bei CrossRef 28.556.113 DOI-Links registriert. Es ist zu erwarten, dass in absehbarer Zeit sämtliche bei kommerziellen Anbietern veröffentlichten wissenschaftlichen Arbeiten über CrossRef nachgewiesen werden können. Alle großen Produzenten wissenschaftlicher Zeitschriften nehmen an CrossRef teil. Was dann noch von deutschen Anbietern (kaum mehr als 5 Prozent des gesamten Zeitschriftenbestandes) fehlt, könnte von dem vom Börsenverein bereitgestellten, aber noch in Entwicklung befindlichen VTO-System nachgewiesen werden (4.3). Das vom Gesetzgeber in Aussicht gestellte Monopol der Informationswirtschaft für die elektronische Dokumentenlieferung auch in Bildung und Wissenschaft (§ 53a UrhG) kann somit sehr bald Realität werden und Bibliotheken entweder auf Printmedien reduzieren oder sie zu kostenaufwändigen Lizenzvereinbarungen mit den Verlagen verpflichten.

¹⁹ Die Syntax eines DOI-Namens besteht aus einem Präfix und, getrennt durch einen Schrägstrich, einem Suffix. Das Präfix beginnt immer mit der Zahl 10, die im Internet der Identifizierung von DOIs dient, und wird, getrennt durch einen Punkt, abgeschlossen durch eine von CrossRef vergebene Nummer für den jeweiligen Informationsanbieter (in der Regel ein Verlag). Für Anordnung und Spezifität des Suffix' ist jeder Informationsanbieter selbst zuständig. Empfohlen wird aber, wie bei URLs, eine hierarchische Anordnung. Der erste Teil des Suffix' enthält also in der Regel den Code für ein Informationsobjekt, zum Beispiel eine Zeitschrift oder auch ein Buch (ersatzweise auch dessen ISBN) oder eine Datenbank. Das kann dann, jeweils getrennt durch einen Punkt, noch weiter spezifiziert werden, zum Beispiel durch Angabe des Erscheinungsjahrs. Von der Idee, die dem DOI zugrunde liegt, können DOIs sich auf beliebig kleine Teile eines Informationsobjekts beziehen, zum Beispiel auf eine spezielle Tabelle, eine spezielle Abbildung oder eine spezielle Literaturreferenz, theoretisch sogar auf jedes einzelne Zeichen. Bei CrossRef werden solche Referenzen und damit Verknüpfungen in die den DOIs zugeordneten Metadaten eingetragen (Richtlinie für Metadaten unter http://Wasserkräfte/02publishers/Mandate_Geldinstitute), die neben vollständigen bibliografischen Angaben auch Verzeichnisse der Abbildungen, Tabellen oder das vollständige Literaturverzeichnis enthalten können.

4 SUCHDIENSTE FÜR (RETRODIGITALISIERTE) BUCHBESTÄNDE

Wie zu Anfang erwähnt, sind hier vor allem die neuen, in der Internet-Umgebung post-professionellen Unternehmen aktiv, die aufgrund ihrer guten Kapitalausstattung in der Lage sind, die Retrodigitalisierung großer Buchbestände zu finanzieren. Dazu gehören neben Google Book Search unter anderem Amazon mit Search Inside!²⁰ sowie Yahoo und Microsoft in der Open Content Alliance. Welche Geschäftsmodelle auf dem Fachinformationsmarkt sich in Zukunft hier entwickeln werden, ist noch offen. Bislang werden die aufgebauten Bestände kaum direkt zur Anzeige genutzt (höchstens bei Copyright-freien Beständen), sondern für Volltextsuchen mit minimaler Textanzeige verwendet. Die Ergebnisse können dann zur Ausleihe oder zum Kauf bei Vermittlungsinstanzen (Bibliotheken, Verlage) führen. Gegenwärtig sieht es danach aus, dass im großen Stil in den Aufbau großer digitalisierter Volltextbestände investiert wird – was damit in der Zukunft mit welchen kommerziellen Verwertungszielen gemacht wird, ist durchaus offen. Unverkennbar ist zudem, dass sich auch die großen Verlage direkt (Beispiel ScienceDirect) oder vermittelt (in Deutschland über die VTO des Börsenvereins) an Such- und Auslieferdiensten auch für elektronische Bücher beteiligen. Daneben entwickelt sich ein neuer Markt für E-Books, die also direkt in elektronischer Form (oft exklusiv elektronisch) produziert und angeboten werden.²¹ Hier macht der alte Vertriebsbegriff kaum noch Sinn. Sie werden zunehmend zu On Demand-Diensten.

4.1 GOOGLE BOOK SEARCH (GBS)

Google begann seinen Volltextnachweis- und Anzeigedienst 2004 mit der Ankündigung (damals noch unter dem Namen Google Print), 15 Millionen Bücher zu digitalisieren, zunächst die Bestände aus fünf amerikanischen und englischen Bibliotheken – Stanford, Michigan, Harvard, New York und Oxford. Es ist die Rede davon, dass Google täglich 50.000 Buchseiten einscannen kann – wie das geht, ist Betriebsgeheimnis, aber es geschieht sicherlich trotz Hochtechnologieinsatz nicht vollautomatisch, sondern vor allem durch eine Vielzahl von Hilfskräften. Anfangs war von 4,5 Milliarden Buchseiten die Rede. Das Projekt sollte einige hundert Millionen US-Dollar kosten und etwa zehn Jahre dauern. Inzwischen sind die Ziele eher noch ehrgeiziger geworden, zumal immer mehr Projektpartner dazukommen, zu denen inzwischen auch deutsche große Bibliotheken wie die Bayerische Staatsbibliothek gehören, die ca. 1 Million lizenzfreie Bücher von Google einscannen und in die Google-Buch-Suche integrieren lässt. Weltweit sind, neben

²⁰ So heißt es auf der Website von Amazon: „Search Inside! ermöglicht die Suche nach Wörtern im Fließtext der Bücher. Mit Search Inside! können Kunden rund um einen Suchbegriff in den Büchern blättern und dort zudem nach weiteren Begriffen suchen. Die Kunden erhalten nur eingeschränkten Einblick in die Bücher und können nicht das gesamte Buch lesen. Dennoch können sie den Kontext, in dem sich ihr Suchbegriff befindet, erkennen und dadurch die richtige Kaufentscheidung treffen. Dies führt dazu, dass Verleger und Autoren mehr Bücher verkaufen! Die Inhalte der Bücher bleiben jedoch geschützt.“ (Amazon 2007).

Verlage müssen entweder ein Freiemplar zur Verfügung stellen, das Amazon digitalisiert, oder aber ein PDF schicken. Weitere Kosten entstehen für den Verlag nicht. Es dauert in der Regel ca. vier Wochen, bis Amazon das neue Werk suchbar nachweist. Der Verlag muss nachweisen, dass er die entsprechenden Verwertungsrechte an dem Werk hat, ggf. auch für diese neue Nutzungsform. Er schließt darüber einen Vertrag mit Amazon.

²¹ Zum Nachweis vgl. Jeré Mias 2007.

den erwähnten, unter anderem Bibliotheken in Madrid, Barcelona, Madison/Wisconsin, Princeton, Austin/Texas sowie die New York Public Library und die Library of Congress dabei. Spektakulär war 2006 der Vertrag mit dem System der kalifornischen Universitäten (UC), an dem allein 100 Bibliotheken beteiligt sind. Bemerkenswert war das vor allem deshalb, weil UC schon vorher über die Open Content Alliance (OCA) – ein Konsortium des nicht-kommerziellen Internet Archive, Yahoo sowie Microsoft – Bestände digitalisieren und verfügbar machen ließ. Google überlässt in der Regel den Partnern eine Kopie der gescannten Werke, beharrt aber überwiegend darauf, dass die Bestände nur über Google-Dienste suchbar sind.

4.2 GBS GOOGLE PRINT

GBS ist nicht nur ein Bibliotheksprogramm (s. 4.4), sondern hat im Google Print-Teil von GBS direkte Vereinbarungen mit den Verlagen getroffen. Hier gibt es nahezu keine Urheberrechtsprobleme, denn in der Regel sollten die Verlage die Nutzungsrechte bei den dann einzuscannenden Informationsobjekten haben. Allerdings haben auch die Autoren ein Wort mitzureden. Im deutschen Urheberrecht § 31 Abs. 4 UrhG gab es bis vor Kurzem eine Regelung, die die Verlage verpflichtete, bei den Autoren eine Erlaubnis für eine zum Zeitpunkt der Erstpublikation unbekannte Nutzungsart einzuholen, auch wenn die Verlage per Vertrag exklusive Nutzungsrechte erworben hatten. Eine unbekannte Nutzungsart liegt bei Büchern, die vor 1995 erschienen sind und die in der Gänze retrodigitalisiert werden sollen, durchaus vor. Nach der jetzt Gesetz werdenden zweiten Urheberrechtsnovellierung ist ein prinzipielles Widerspruchsrecht nicht mehr vorgesehen. Bei Werken, die ab 1.1.1965 erschienen sind und für die exklusive Nutzungsrechte eingeräumt wurden, gelten die Nutzungsrechte auch für Nutzungsarten, die beim Erscheinen unbekannt waren, als eingeräumt, es sei denn, der Autor legt dagegen explizit innerhalb eines Jahres nach Gültigwerden der zweiten Urheberrechtsanpassung Widerspruch ein (zu erwarten mit Veröffentlichung im Bundesanzeiger noch 2007). Hat der Verlag schon die neue Nutzungsart geplant und den Autor darüber informiert, so erlischt die Widerspruchsfrist sogar schon nach drei Monaten. Gibt es also kaum Urheberrechtsprobleme mit den Verlagen, so könnten doch Autoren durchaus ihre Rechte anmelden. Wenn deutsche Autoren nicht mit der Digitalisierung ihrer Werke einverstanden sind, sollten sie ab sofort Google einen Brief schreiben und diese untersagen. Die jetzt vorgesehene Drei-Monate-Frist wird kaum Anwendung finden, denn Google wird wohl nicht alle die Autoren anschreiben wollen oder können, deren Bücher zur Digitalisierung anstehen.

Den Verlagsinteressen trägt Google Print dadurch Rechnung, dass Nutzer in der Regel nur sehr kurze Abschnitte (Snippets), zum Beispiel Sätze mit dem Suchbegriff im Kontext, oder sogar nur bibliografische Information einsehen können. Falls Verlage es erlaubt haben, können mehrere Seiten eingesehen werden. Registrierte Nutzer können sich noch weitere Seiten anschauen – so lange, bis ihr festgelegtes Tageskontingent

erschöpft ist. Durch Anwendung von DRM-Techniken soll der Ausdruck und das Abspeichern von diesen Snippets verhindert werden. Rechtsfreie Texte können vollständig eingesehen werden, auch andere Texte, wenn Verlage oder Autoren das explizit erlauben. In jedem Fall wird auf Verlage bzw. Online-Buchhandlungen verwiesen, von denen die Texte erworben werden können. Ebenfalls werden Hinweise auf Bibliotheken gegeben, bei denen das betreffende Werk ausgeliehen werden kann. Neben bibliografischer Information werden noch Metainformationen (textkennzeichnende Key Words oder Key Phrases) bereitgestellt. Nutzer können Reviews schreiben und Bewertungen hinterlegen, ebenso ausgewählte Titel in ihre eigene elektronische Bibliothek einfügen, die sie auch anderen Nutzern zur Einsicht geben können. Damit teilt Google Print durchaus auch einige Merkmale mit sozialen Diensten, die Gegenstand des sogenannten Web-2.0-Paradigmas sind, auf das sich auch Google zunehmend einlässt.

Das Partner-Programm ist also wie viele anderen Anzeigendienste auch in erster Linie als ein Marketing-Mittel einzuschätzen. Wie erfolgreich Google Print ist, das heißt wie viele Verlage ihre Materialien über Verträge bereitstellen werden, ist schwierig einzuschätzen. Verlage arbeiten ja selbst an solchen Volltextsuchsystemen, so in Deutschland der Börsenverein des Deutschen Buchhandels mit dem VTO-Projekt, auf das ich nachfolgend eingehe, obgleich derzeit noch nicht klar ist, wie weit es sich wirklich durchsetzen wird.

4.3 VOLLTEXTSUCHE-ONLINE (VTO/LIBREKA)

Das vom Börsenverein des Deutschen Buchhandels initiierte Projekt Volltextsuche-Online (VTO, vor Kurzem umbenannt in „libreka“) bezieht sich auf digitale Volltextdateien von deutschsprachigen Büchern²². „Mit Volltextsuche-Online bereitet der deutsche Buchhandel die Grundlage dafür, dass Buchinhalte in Zukunft in elektronischer Form recherchierbar, verfügbar und handelbar werden“ (aus der Mitte Februar 2007 zunächst für Verlage freigeschalteten Website).²³

Offenbar ist das Projekt immer noch nicht aus den Startlöchern gekommen; jedenfalls ist den verschiedenen öffentlichen Ankündigungen noch kein öffentlich websichtbares Ergebnis zu entnehmen. Angezeigt werden durch VTO nicht die vollständigen Volltexte, sondern einige Seiten aus dem elektronischen Volltext. Jeder Verlag kann flexibel festlegen, wie viele Seiten er angezeigt haben will. Durch die vertragsmäßig abgesicherte Beteiligung der Verlage als Rechteinhaber ihrer Publikationen gibt es nach den Vorstellungen des Börsenvereins keine Urheberrechtsprobleme (s. aber Abschnitt 4.2). Es ist auch nicht zu erwarten, dass VTO gegen die Normvorschriften bezüglich der unbekanntenen Nutzungsarten verstoßen wird. Inwieweit das Angebot flächendeckend angenommen wird, bleibt angesichts der Gebühr, die der Börsenverein beim Verlag erhebt – pro eingestelltes Buch sind 17 Euro für zwölf Monate vorgesehen –, abzuwarten. Zudem bleiben die Kosten zur Digitalisierung den Verlagen überlassen, während bei den kommerziellen

²² Börsenverein des deutschen Buchhandels 2007.

²³ <http://info.libreka.de/>.

Diensten wie Google oder Amazon lediglich die gedruckten Vorlagen bereitgestellt werden müssen. Bei VTO müssen die Verlage zudem PDF-Dateien und die entsprechenden Metadaten bereitstellen (neben den bibliografischen Daten auch die Spezifikation des Umfangs der Zugriffsrechte). Auch die Qualitätskontrolle, zum Beispiel über die richtige Anzeige, bleibt bei den Verlagen. Zum Start hatten 20 Verlage den Nutzungsvertrag unterschrieben, für weitere 160 Verlage lagen Absichtserklärungen vor.²⁴ Um Endkunden die Nutzung zu erleichtern, ist – unbeschadet der Klage des Börsenvereins gegen Google wegen der Recherche zu deutschsprachigen urhebergeschützten Werken in Google Book Search²⁵ – eine Zusammenarbeit mit den Standardsuchmaschinen vorgesehen. Diese sollen dann auf eine eigene Indexierung verzichten, wenn das über VTO schon geschieht. Es ist allerdings geplant, bei den Suchmaschinen nur die Titel als Ergebnis anzeigen zu lassen, nicht die Detailanzeige selbst wie bei VTO zu erlauben. Auch hier bleibt abzuwarten, inwieweit dem internationalen Markt die Bedingungen vorgeschrieben werden können.

4.4 GBS LIBRARY PROGRAM

Urheberrechtlich problematischer sieht es beim GBS Library Program aus, bei dem, wie unter 3.1 erwähnt, im großen Stil gesamte Bibliotheksbestände ohne Einholen der Erlaubnis der Rechteinhaber (seien es Autoren oder Verlage) eingescannt und suchbar gemacht werden. Anders als die schon erwähnte Open Content Alliance, die sich nicht zuletzt durch die Teilnahme des von Brewster Kahle gegründeten Internet Archive auch als Open Source/Access-Alternative zu Google versteht, beschränkt sich Google nicht prinzipiell auf Copyright-freie Objekte, sondern will – es sei denn, die Verträge sehen, wie bei der Bayerischen Staatsbibliothek, anderes vor – zunächst einmal alles aus den Beständen der jeweiligen Bibliothek einscannen, um es dann herauszunehmen, wenn Urheberrechtsbedenken geäußert werden (Opt out-Prinzip).

Hier wird die Rechtslage durchaus kontrovers eingeschätzt. Von Beginn an sah sich Google Klagen vor allem verschiedener Verlage ausgesetzt. An prominenter Stelle in Deutschland ist hier die Wissenschaftliche Buchgesellschaft Darmstadt zu erwähnen (im Mai 2006 vor dem Landgericht München), unterstützt vom Börsenverein des Deutschen Buchhandels, der selbst mit seinem VTO- bzw. libreka-Projekt Interesse an der Anzeige von digitalen Volltexten seiner Verlagsmitglieder hat. Auf der anderen Seite bekam Google zeitweise durchaus theoretische ökonomische und juristische Unterstützung. Eine Opt in-Variante, nämlich vorab zu klären, wie die Rechtssituation aussieht, sei aus wirtschaftlichen Gründen nicht zu realisieren²⁶, da aufwändige Recherchen nach dem jeweiligen Rechteinhaber anfallen würden. Die dann entstehenden Transaktionskosten würden das gesamte Geschäftsmodell zugrunde richten. Juristisch seien die Vorwürfe nicht haltbar, weil die Indexierung der gescannten Werke, die ja dem Dienst zugrunde liegt, weder (in der deutschen Urheberrechtssprache) Wiedergabe, noch öffentliche Zugänglichmachung

²⁴ Vgl. bildungsklick.de 2007.

²⁵ Die Klage wurde zunächst vom Landgericht Hamburg abgewiesen, da wegen der geringen Anzeige von wenigen Zeilen als Ergebnis einer Suche in Google Book Search eine Urheberrechtsverletzung kaum vorliege. Problematisch ist zudem die Zuständigkeit eines deutschen Gerichts für den US-amerikanischen Dienst. In den USA scheint die Nutzung mit dem Fair-use-Prinzip verträglich zu sein. Vgl. Die Zeit 2007.

²⁶ Vgl. dazu den Bericht von Monika Ermert: „Google Book Search: Opt-in oder Opt-out?“, in: Heise 2006.

sei, sondern lediglich (vergleichbar den Suchmaschinen, deren Indexerstellung ja auch nicht als Copyright-Verletzung kritisiert werde) eine Art Katalog oder eben ein Suchindex. Daher sei auch keine Konkurrenzsituation zu den Vermarktungsinteressen der Rechteinhaber gegeben. Die Anzeige der Snippets, die ja als Wiedergabe interpretiert werden kann, sei durch das US-amerikanische Fair use-Prinzip gedeckt, da die Anzeige selber auch keinem direkten kommerziellen Zweck diene. Schließlich werden die Snippets oder andere angezeigte Teile nicht verkauft oder geleast. Auf der anderen Seite, bei der Kritik am Opt in-Modell, wird vorgetragen, dass es den Autoren bzw. Rechteinhabern nicht zuzumuten sei, selber herauszufinden, ob ihre Werke in GBS oder einem der vielen anderen zu erwartenden Volltextdiensten eingescannt und suchbar gemacht worden sind.

Das Ganze ist ein in den letzten Jahren intensiv diskutiertes Problem, wie denn mit verwaisten Werken (Orphan Works) umzugehen ist. Gemeint sind damit Werke, deren Rechteinhaber unbekannt sind bzw., wenn bekannt, nicht ausfindig gemacht werden können, oder es sind Werke, bei denen es schwierig oder unmöglich ist, die genaue Dauer des Urheberrechts festzustellen. In den USA werden zu den Orphan Works auch vergriffene Werke gezählt. Daher kann der Anteil der Orphan Works an den gesamten in den USA publizierten Büchern in den Jahren ab 1960 zwischen 30 und 50 Prozent betragen, wobei der Anteil des Commercial Abandonment bei Weitem am größten ist. Das Ausmaß verwaister Werke, und damit verbunden die rechtlichen Unsicherheiten, sind bei anderen medialen Objekten wie Filmen eher noch größer und auch von größerer wirtschaftlicher Relevanz. Das Problem der Rechtsunsicherheit bei verwaisten Werken ist also in quantitativer Hinsicht nicht unerheblich und kann wegen des damit verbundenen Aufwandes (vor allem bei Mehrfachautoren) eigentlich nicht durch individuelle Recherche und Einholen der Nutzungs- oder Scan-Erlaubnis gelöst werden. Bibliotheken schrecken angesichts der mit verwaisten Werken verbundenen Probleme in der Regel vor einer Digitalisierung und Bereitstellung dieser Werke zurück, sodass ein wichtiges Kulturerbe der digitalen Nutzung verschlossen bleibt.

Hier spielt Google – so merkwürdig sich das auch aus der Sicht der Kritiker an Googles kommerziellem Verwertungsmodell und an seiner bedrohlichen Monopolstellung bei den Suchmaschinen anhören mag – mit seiner Herausforderung an die existierenden Copyright-Vorschriften eine wichtige Rolle. Nicht dass Google durch seine Geschäftsmodelle bezüglich GBS die Gültigkeit bestehender Copyright-Regelungen verneinte; faktisch zielen diese Modelle jedoch auf eine Infragestellung von Regelungen ab, die in den letzten Jahren immer mehr dem Schutz vor der direkten kommerziellen Verwertung von Informationsprodukten jeder Art gedient haben. Aus der Google-Sicht ist dies heute eher innovationsbehindernd als innovationsfördernd. Vielleicht wird es einmal zu den historischen Verdiensten von Google gerechnet werden, dass es unter dem Motto, das Wissen der Welt allen frei verfügbar zu machen, keinen Sinn macht, in elektronischen Umgebungen mit den Informationsprodukten, den Werken selber, Gewinn machen zu wollen, und es daher auch keinen Sinn macht, den kommerziellen Handel mit Wissen und Information über das Urheberrecht schützen zu wollen.

5 EINIGE TRENDS

Der Fachinformationsmarkt gerät derzeit stärker noch als bislang unter die Kontrolle der Informationswirtschaft. Bibliotheksdienste können da nur teilweise mithalten – aus organisatorischen, technisch-informationsmethodischen und rechtlichen Gründen. Die historisch gewachsenen, zumal in Deutschland stark dezentral organisierten bibliothekarischen Dienste tun sich bei der Vielzahl von Ordnungssystemen, formalen und inhaltlichen Erschließungsformen sowie heterogenen Formaten nach wie vor schwer, Nutzern ein einheitliches Angebot zu machen, obgleich unverkennbar ist, dass hier enorme Fortschritte gemacht worden sind. Angebote wie die von subito, Vascoda,²⁷ der Karlsruher digitalen Bibliothek oder die zahlreichen von der DFG geförderten und zunehmend vernetzten virtuellen Bibliotheken²⁸ – um nur diese exemplarisch zu nennen – haben für Nutzer größere Transparenz entstehen lassen. Organisatorisch ist der öffentliche Informationsmarkt also auf einem guten Weg (gewesen).

Weniger gelungen ist die technische und informationsmethodische Umsetzung. Auch in Bildung und Wissenschaft erwarten Kunden den gleichen hohen Standard im Benutzer-Interface, in der Verfügbarkeit und Ausfallsicherheit sowie in der Bereitstellung informationeller Mehrwerte. Dies wird aber in elektronischen Umgebungen immer kostenaufwändiger. Angesichts der systematischen Unterfinanzierung des öffentlichen Bibliotheksbereichs können zwar über Projektfinanzierung (insbesondere durch DFG und BMBF) einige Dienste mit guten Konzepten auf den Weg gebracht werden. Die dauerhafte institutionelle Absicherung bleibt aber dann in der Regel aus, zumal, wie oben erwähnt, die institutionelle Infrastruktur in Deutschland von der Politik weitgehend aufgelöst wurde. Die meisten so entstandenen Dienste können mit den auf dem Markt etablierten Standards, wie exemplarisch bei ScienceDirect/Scirus oder CrossRef realisiert, kaum konkurrieren.

Etwas besser sieht es bei den in den 70er und 80er Jahren des vergangenen Jahrhunderts auch aus der öffentlichen Förderung entstandenen Fachinformationszentren aus, die ihre Qualität und Sicherheit aber nur halten können, weil sie durch den Druck ihrer öffentlichen Geldgeber gezwungen sind, kostendeckend zu arbeiten – was bei Zentren wie dem FIZ Karlsruhe, FIZ Technik oder FIZ Chemie ja auch mit erstaunlichen Kostendeckungsgraden gelungen ist, ganz zu schweigen von Angeboten wie JURIS, die ja gänzlich in den Markt entlassen worden sind. Ob dies Bildung und Wissenschaft entgegen kommt, kann allerdings bezweifelt werden.

²⁷ Vascoda (www.vascoda.de), entstanden durch gemeinsame Förderung des Bundesministeriums für Bildung und Forschung (BMBF) und der Deutschen Forschungsgemeinschaft (DFG), ist ein Internet-Portal für Wissenschaftsinformation. Der informationelle Mehrwert besteht wesentlich in einer einheitlichen Oberfläche und im Nachweis von Wissensobjekten auch aus Ressourcen, die in der Regel für Internet-Suchmaschinen unsichtbar sind. Vascoda ist mit öffentlichen und kommerziellen Nachweissystemen und Dokumentenlieferdiensten verknüpft, sodass (ggf. gegen Entgelt) die Volltexte (elektronische Zeitschriftenartikel) geliefert werden können.

²⁸ Der Karlsruher Virtuelle Katalog (KVK) ist ein Meta-Katalog zum Nachweis von mehr als 500 Millionen Büchern und Zeitschriften in weltweiten Bibliotheks(verbund)katalogen und Buchhandelsverzeichnissen (zum Beispiel auch amazon.de, libri.de oder ZVAB). vgl. www.ubka.uni-karlsruhe.de/hylib/virtueller_katalog.html.

Die öffentliche Hand, gemeint sind hier wegen ihrer institutionellen Zuständigkeit die Länder, hat unübersehbar nicht den Willen gezeigt bzw. nicht in ausreichendem Maße die Kraft aufgebracht, die klassischen Informationsvermittlungseinrichtungen (insbesondere die Bibliotheken) finanziell und technisch so auszustatten, dass sie zum einen mit den im Internet selbstverständlich gewordenen Komfortstandards und zum anderen mit den kommerziellen, zunehmend auf die Endnutzer ausgerichteten Diensten mithalten könnten. Der Verdacht ist nicht ganz von der Hand zu weisen, dass, wenn nicht bewusste politische Absicht, so doch eine Art politische Invisible Hand die Prozesse, die wir mit Begriffen wie „Dismediatisierung“, „Kommerzialisierung“, „Retaildienste“ und „informationelle Mehrwerte“ beschrieben haben, fort von der öffentlichen Informationsversorgung hin zur kommerziellen Informationswirtschaft befördert hat.

Diese Vermutung wird nicht zuletzt dadurch bestärkt, dass die urheberrechtlichen Regulierungen der letzten Jahre es den Bibliotheken immer schwerer gemacht haben, ihre Positionen gegenüber der Informationswirtschaft zu bewahren oder gar auszubauen. Vielleicht waren sie anfänglich anders intendiert, denn vor allem die letzte Urheberrechtsnovellierung (im Zweiten Korb) hat ja durchaus bibliotheksbezogene Schranken, also Ausnahmen von den exklusiven Urheber- und Verwerter-/Nutzungsrechten eingeführt – die sich dann im Prozess der Aushandlung aber eher ins Gegenteil verkehrt, also der Fortschreibung des kommerziellen Primats gedient haben.

Nach den jetzt im Zweiten Korb getroffenen urheberrechtlichen Regelungen des elektronischen Dokumentenversands zur Informationsversorgung in Bildung und Wissenschaft durch Bibliotheken (§ 53a UrhG) dürfen Bibliotheken keine Informationsobjekte in elektronischer Form mehr an ihre Kunden ausliefern (und da auch nur für wissenschaftliche Zwecke), wenn es für Bibliotheken „offensichtlich“ erkennbar ist, dass der Markt ein entsprechendes Angebot bereitstellt.²⁹ Über CrossRef, vielleicht auch über Vascoda, dürfte diese Bedingung der „offensichtlichen“ Erkennbarkeit für so gut wie alle Objekte als erfüllt anzusehen sein, zumindest für die in der Wissenschaft zentralen Artikel in wissenschaftlichen Zeitschriften.³⁰

Nun ist die Regulierung über das Urheberrecht nicht der einzige Weg, Zugänge zur wissenschaftlichen Information zu eröffnen oder zu verknappen. Lizenzgebühren, also vertragliche Regelungen, können das Urheberrecht erweitern, modifizieren oder korrigieren – so, wie es jetzt auch subito notgedrungen macht. Auch die DFG hat enorme Anstrengungen gemacht, um über Nationallizenzen (in Erweiterung des Konzepts

²⁹ Die zweite Bedingung ist, dass das Angebot zu angemessenen Bedingungen erfolgt. Hier muss abgewartet werden, inwieweit sich bei der Entwicklung des kommerziellen wissenschaftlichen Informationsmarktes als elektronisches Medium zu einem Massenmarkt die bislang eher noch restriktiven Preise für einen Artikel (durchschnittlich ca. 30 Euro) auf „angemessen“ reduzieren werden. Die Bedingung der „Angemessenheit“ kann dann aber auch, wie es der deutsche Bibliotheksverband anvisiert, durch umfassende Lizenzvereinbarungen erfüllt werden. Auf den wissenschaftlichen Informationsmärkten sind also durchaus spannende Entwicklungen und Auseinandersetzungen zu erwarten.

³⁰ Mitte September 2007 nahmen weltweit schon 1.260 Bibliotheken am CrossRef-System teil, darunter auch viele große deutsche Bibliotheken, bislang eher zur Überprüfung bzw. Erweiterung ihrer Online-Kataloge. In Zukunft wird diese Beteiligung sicherlich vermehrt in Anspruch genommen, um die Berechtigung eines Dokumentensands zu verifizieren (bzw. immer mehr diese zu verneinen, da das Angebot über CrossRef ersichtlich ist).

der Sondersammelgebiete) die Versorgung mit elektronischer Fachinformation an den Hochschulen und Forschungseinrichtungen über die Bibliotheken sicherzustellen.³¹ Der Zugriff selbst soll für Studierende und Forscher kostenlos sein, und zwar nicht nur im hausinternen Netzwerk und an speziellen Arbeitsplätzen (wie jetzt im § 52b UrhG gefordert), sondern er soll auch „ohne Einschränkungen über Weitverkehrsnetze den überwiegend öffentlich geförderten Hochschulen und Forschungseinrichtungen in Deutschland bzw. wissenschaftlichen Nutzern mit ständigem Wohnsitz in Deutschland zugänglich“ gemacht werden.³² Für diese private Nutzung sind eine persönliche Registrierung und Passwortvergabe erforderlich. Bei einer institutionellen Nutzung werden die jeweiligen IP-Nummernkreise freigeschaltet.

Bibliotheken gehen auch sonst zunehmend Private Public-Kooperationen mit der Verlagswirtschaft ein – freiwillig oder gezwungenermaßen. Niemand hat die Bayerische Staatsbibliothek gezwungen, sich auf das Google Library Program einzulassen. Allerdings hätte sie die Retrodigitalisierung ihrer (bislang urheberrechtsfreien) Bestände wohl kaum aus eigenen Mitteln finanzieren können. Problematisch werden könnte dies, wenn dabei bibliothekarische Qualitätsstandards bezüglich der Metainformationen verloren gingen. Dramatisch würde es, wenn die Googles sich über die Verträge weitgehend die Verwertungsrechte (auch für noch unbekanntes Nutzungsarten) sichern könnten, sodass den Bibliotheken nur die Archivierungsleistung und vielleicht noch die lokale Nutzung entsprechend § 52b UrhG übrig bliebe. Es ist nicht ausgeschlossen, dass Bibliotheken, wie heute schon bei den wissenschaftlichen Aufsätzen, dann auch für die elektronische Nutzung ihrer eigenen Buchbestände Lizenzverhandlungen mit den durch die Retrodigitalisierung entstandenen neuen Rechteinhabern werden führen müssen.

Andere Kooperationen wie bei Onleihe (s. Abschnitt 3.3) setzen auf Mehrwertleistungen, die sie alleine nicht erbringen könnten. Auch hier reguliert das Urheberrecht, über §§ 53 und 95a, mit der Folge, dass in den Bibliotheks-, Bildungs- und Wissenschaftsbereichen (im kommerziellen Bereich sowieso) rechtlich geschützte DRM-Techniken in den Dokumentlieferdiensten eingesetzt werden – zu einer Zeit, in der die internationale Musikindustrie zunehmend von DRM-Techniken absieht. Wollen Bibliotheken in der Zukunft nicht nur für Druck- und Fax-Ausleihen zuständig sein, werden sie mit den kommerziellen Rechteinhabern Lizenzvereinbarungen aushandeln müssen, deren Nutzung zum einen DRM-kontrolliert sein wird und die zum anderen höhere Informationskosten für die Nutzer auch in Bildung und Wissenschaft mit sich bringen werden, die dann entweder von den Ländern als Träger der Hochschulen und Bibliotheken oder von den Endnutzern selbst, in erster Linie von den Auszubildenden, zu tragen sind.

³¹ www.nationallizenzen.de/ueber-nationallizenzen [Stand: 10.10.2007]. Bisher haben größere Institutionen entsprechende Lizenzverträge abgeschlossen: Bayerische Staatsbibliothek, Staatsbibliothek zu Berlin, Universitätsbibliothek Frankfurt, Technische Informationsbibliothek Hannover, GESIS – Informationszentrum Sozialwissenschaften, Universitäts- und Stadtbibliothek Köln, Staats- und Universitätsbibliothek Göttingen, Deutsche Zentralbibliothek für Wirtschaftswissenschaften sowie die Deutsche Zentralbibliothek für Medizin.

³² Ebenda.

Viele der durch die Private Public-Partnerships sich abzeichnenden Entwicklungen – gemeint sind hier, wie immer in diesem Beitrag, die Entwicklungen für Bildung und Wissenschaft – müssen auf der Nutzerseite höchst kritisch gesehen werden. Auch hier werden Informationsleistungen immer mehr als elektronische Dienste nachgefragt. Informationsnutzer erwarten, dass ihnen die Bibliotheken die Information in elektronischer Form genauso gebührenfrei und nutzerfreundlich anbieten, wie sie es in der Vergangenheit mit analogen Materialien und guter, individueller Benutzerbetreuung geboten haben. Das werden sie aber in Zukunft immer weniger tun können und auch nicht dürfen.

In Bildung und Wissenschaft wird es schwierig zu vermitteln sein, dass auf der einen Seite der alltägliche Informationsbedarf problemlos und gebührenfrei aus dem Internet gedeckt werden kann, während auf der anderen Seite die Situation für die Deckung des für die Gesellschaft weitaus wichtigeren Informationsbedarfs in der Ausbildung und der Forschung immer problematischer wird: Entweder – wenn es denn weiter gebührenfrei sein soll – sollen nur analog präsentierte und ausgelieferte Wissensobjekte verwendet werden dürfen, oder es müssen für die elektronische Lieferung erheblich höhere Gebühren entrichtet werden. Elektronische Informationsdienstleistungen, an sich dafür konzipiert, den Umgang mit Wissen und Information unkomplizierter und kostengünstiger (oder sogar gebührenfrei) zu machen, werden stärker (DRM-) kontrolliert und durch Lizenzvereinbarungen mit den kommerziellen Anbietern teurer – für Auszubildende vermutlich unerschwinglich.

Dabei werden die Informationskosten in Zukunft vermutlich gar nicht einmal mehr die unerschwinglichsten sein. Die Dienste der Googles sind gebührenfrei; aber auch wenn sie dies in elektronischen Räumen zukünftig bleiben sollten, so muss immer gezahlt werden, wenn auch in anderen ‚Währungen‘ wie der Offenlegung der Nutzungsdaten, des Nutzungsverhaltens, der Kooperations- und Kollaborationsformen in den von den Googles ebenfalls betriebenen sozialen Diensten, der Vorlieben und Abneigungen – und was auch immer noch freigegeben, gespeichert und daraufhin für kommerzielle Zwecke verwendet werden wird. Google (als Metapher für all die neuen Dienste, von denen die hier in den Vordergrund gestellten elektronischen Volltext- oder Buchdienste nur ein Muster sind) ist das „Opium der Informationsgesellschaft“: süß, gebührenfrei, wenn auch nicht kostenlos, verführerisch, süchtig und abhängig machend. Kommt dazu noch die sich eher intensivierende Monopolisierung der elektronischen Märkte, dann zeichnet sich hier öffentlicher Handlungsbedarf ab: Die Googles werden entflochten, reguliert werden müssen.

Es gibt jedoch keinen Trend ohne Gegentrend. Hier kommen zunächst und mit dem Vorbehalt des Opium-Merkmals noch einmal die neuen post-professionellen Akteure ins Spiel, exemplarisch in diesem Beitrag mit GBS vorgestellt. Google ist auch schon mit seiner allgemeinen Suchmaschine, ergänzt durch Google Scholar, zu einem wichtigen Nachweisdienst auch in der Wissenschaft geworden. Da Wissenschaftler ihre Publika-

tionen zunehmend heute schon entweder als Direkt- und damit Primärpublikation, oder als Sekundärpublikation mit Erlaubnis der Verlage (so Elsevier) ins Web stellen, können, vermittelt über die Google-Suchleistung, viele wissenschaftliche Volltexte frei (das heißt DRM- und gebührenfrei) eingesehen werden. Welche Geschäftsmodelle sich mit Google Book Search und den anderen Suchdiensten für (retrodigitalisierte) Buchbestände entwickeln werden, unter denen dann nicht nur Snippets, sondern auch ganze, urheberrechtlich geschützte Volltexte eingesehen (und weiter genutzt) werden können, ist allerdings durchaus offen.

Der eigentliche Gegentrend kann nur direkt aus der Wissenschaft selbst kommen. Dies ist ja in der Tat mit den beiden Zweigen der Open Access-Publikation der Fall. „Open Access“ bedeutet in jedem Fall die kostenfreie und auch unrestringierte Nutzung von Wissensobjekten jeder Art, nicht nur, um sie nutzend wahrzunehmen, sondern auch je nach Bedürfnis weiterverarbeiten oder sogar modifizieren zu dürfen. Dies kann bekanntlich auf dem Weg der speziellen Open Access-Zeitschriften (dem „goldenen Weg“) oder über den Weg der Open Access Repositories (dem „grünen Weg“) geschehen, in denen die bereits publizierten Werke zweitverwertet werden, also entweder direkt oder mit einem geringen Zeitembargo versehen (in der Diskussion sind sechs Monate) in die Repositories aufgenommen und verfügbar gemacht werden. Für beide Wege gilt natürlich, dass die kostenlose Nutzung nicht mit einer kostenlosen Produktion, Speicherung und öffentlichen Bereitstellung gleichbedeutend ist.

Open Access wird sich in absehbarer Zeit auch auf Volltexte, die bislang in Buchform entstanden sind, erstrecken. Entsprechend werden sich auch hier Geschäfts- und Organisationsmodelle zur Sicherung der kostenlosen Nutzung in Bildung und Wissenschaft entwickeln. Aber genauso, wie bei Erwerb und Erschließung der Bücher und Zeitschriften durch Bibliotheken die öffentliche Hand für die Kosten aufkommen musste, so werden auch die Open Access-Modelle finanziert werden müssen.

Es ist abzusehen, dass die jetzigen kommerziellen Volltext-/Buch-Dienste der Googles auch in die Bereitstellung von Repositories einsteigen werden, so, wie sie auch jetzt schon kostenlos die gesamte E-Mail-Verwaltung für Universitäten übernehmen, wenn das entsprechende Angebot akzeptiert wird. Und genauso, wie sich für die Zeitschriften die Verwertungsmodelle entwickeln, wird das auch für die Volltextobjekte in Form der Bücher der Fall werden.

Im Interesse von Bildung und Wissenschaft kann man nur hoffen, dass die bisherigen Financier von Hochschulen und Bibliotheken in der gebührenfreien Nutzung durch die neuen Akteure nicht eine willkommene Gelegenheit sehen, die bisherigen hohen Informationskosten zum Unterhalt der Bibliotheken und anderer öffentlicher Informationsvermittlungseinrichtungen tendenziell gegen Null herunterzufahren. Bildung und Wissenschaft würden dann von den kommerziellen Interessen noch abhängiger und müssten mit ganz anderer und weitaus einschränkenderer „Währung“ den Rückgriff auf das bislang publizierte Wissen bezahlen.

Auf den wissenschaftlichen Fachkommunikationsmärkten ist so gut wie alles im Fluss. Die Opium-artige Wirkung der Googles darf nicht darüber hinwegtäuschen, dass die Informationsversorgung für Bildung und Wissenschaft nach wie vor, in welcher Form auch immer, eine öffentliche und weiter kostenintensive Aufgabe ist, die nicht gänzlich an den Markt abgegeben werden kann. Das schließt aber keineswegs aus, dass aus dem Markt heraus neue, elektronischen Umgebungen angemessene Geschäftsmodelle entwickelt werden, die den Verlagen sowie den neuen Akteuren – unter Anerkennung des Primats des freien (kostenlosen und unbeschränkten) Zugriffs auf Wissen und Information, vor allem des mit öffentlichen Mitteln produzierten Wissens – auch eine kommerziell attraktive Nutzung dieses Wissens erlaubt. Es wird aber auch von der Bereitschaft der Wissenschaftler selbst abhängen, inwieweit sie bereit sind, die Open Access-Publikationsformen anzunehmen.

6 LITERATUR

Amazon 2007

URL: <http://www.amazon.de/gp/help/customer/display.html/303-7122296-5020217?ie=UTF8&nodeId=14209981> [Stand: 10.10.2007].

bibliothek digital 2007

URL: <http://www.bibliothek-digital.net/index.php?id=4> [Stand: 10.10.2007].

Bildungsklick.de 2007

URL: <http://bildungsklick.de/pm/50312/plattform-volltextsuche-online-2007-startklar-fuer-verlage> [Stand: 10.10.2007].

Börsenverein des deutschen Buchhandels 2007

URL: <http://www.boersenverein.de/de/108603?template> [Stand: 10.10.2007].

Die Zeit 2007

www.zeit.de/news/artikel/2006/10/04/76062.xml [Stand: 10.10.2007].

DiViBib 2007

URL: <http://www.divibib.com/18.0.html> [Stand: 10.10.2007].

Elsevier 2007

URL: http://www.elsevier.com/wps/find/authored_newsitem.cws_home/company-news05_00077 [Stand: 10.10.2007].

Heise 2006

URL: <http://www.heise.de/newsticker/meldung/70139> [Stand: 10.10.2007].

Heise 2007

URL: <http://www.heise.de/tp/r4/artikel/26/26064/1.html> [Stand: 10.10.2007].

Jeré Mias 2007

URL: <http://www.jere-mias.de/docu/epublish.html> [Stand: 10.10.2007].

Scirus 2007

URL: <http://www.scirus.com/srsapp/aboutus> [Stand: 10.10.2007].

Springer 2007a

URL: <http://www.springer.com/dal/home/open+choice?SGWID=1-40359-12-115382-0> [Stand: 10.10.2007].

Springer 2007b

URL: http://www.springer.com/dal/home/open+choice?SGWID=1-40359-12-161193-0&teaserId=55557&CENTER_ID=115393 [Stand: 10.10.2007].

Stadt Würzburg 2007

URL: http://www.wuerzburg.de/de/kultur-bildung/stadtbuecherei/vr/11109.Was_ist_die_Virtuelle_Stadtbuecherei_Wuerzburg.html [Stand: 10.10.2007].

STM 2007

URL: <http://www.stm-assoc.org> [Stand: 10.10.2007].

subito 2006

URL: <http://www.subito-doc.de/cms/filedatabase/files/Presseerklaerung%20deutsch.pdf> [Stand: 10.10.2007].

subito 2007

URL: <http://www.subito-doc.de/index.php?lang=de&mod=page&pid=Urheberrecht&PHPSESSID=2b9e09f84e9d5bc718d5decfd730e2e8> [Stand: 10.10.2007].

> AUF DEM WEG IN DIE GOOGLE-GESELLSCHAFT?

RUDI SCHMIEDE

1 EINLEITUNG

Ich möchte im Folgenden mit der Frage beginnen, worum es bei dem Thema „Google-Gesellschaft“ geht: Wir haben es mit neuen Formen des Umgangs mit dem Wissen der Welt zu tun. Ich werde dann kurz erläutern, worin der ökonomische, gesellschaftliche und politische Hintergrund der diskutierten Entwicklung besteht, und dabei auf den informationellen Kapitalismus, die Herausbildung der Netzwerkgesellschaft und die Entwicklung von Google eingehen. Anschließend möchte ich kurz charakterisieren, was die wesentlichen Momente der gegenwärtigen Informatisierung sind und in welcher Weise dieser Prozess mit neuen Wissensanforderungen verbunden ist. Abschließend werde ich auf die Herausbildung neuer Kompetenzen eingehen und ihre Chancen und Gefahren diskutieren.

2 WURUM ES GEHT: DER UMGANG MIT DEM WISSEN DER WELT

Google (das Wort wird im Folgenden als stellvertretende Benennung für die ganze Familie neuerer Suchmaschinen gebraucht) stellt eine neuartige und zentrale Schnittstelle zwischen Mensch und Information dar. Erstmals werden im globalen Maßstab sowohl sortierte als auch nach Bedeutungskriterien gegliederte Wissensbestände zugänglich, und dies vom einzelnen Arbeitsplatz aus. Mit diesem neuen Zugang zum Wissen der Welt sind jedoch auch weitere Dimensionen des „Digital Divide“ verbunden: Die Spaltung in digitales und nicht-digitalisiertes Wissen entscheidet über die Zugangsmöglichkeiten an dieser Schnittstelle. Google steht so einerseits für einen erweiterten Zugang zum Wissen der Welt, auf der anderen Seite aber markiert die Suchmaschine auch dessen limitierende Definition. Google prägt also einen ganz wichtigen Bereich unserer heutigen Wahrnehmung der Realität; man kann sogar sagen, dass es einen Beitrag zur Konstitution dieser Realität leistet, die ja immer durch unsere Wahrnehmung geprägt ist. Dies ist sicherlich die Grundlage dafür, dass sich heutzutage eine neue Ökonomie der Aufmerksamkeit beziehungsweise der Aufmerksamkeitsvorteile herausgebildet hat.¹ Die extrem hohe Bewertung des Unternehmens Google durch die Börse lässt sich nur verstehen, wenn man sich klar macht, dass es sich hier um ein neues Geschäftsmodell des Handels

¹ Vgl. Franck 2003.

mit Meta-Informationen handelt. Google verfügt über riesige Mengen an akkumulierten Informationen über den Umgang von Personen mit Informationen, die von vielen Marktteilnehmern als wichtiges und wertvolles Gut betrachtet werden.

3 INFORMATIONELLER KAPITALISMUS UND GOOGLE

Der von Manuel Castells so bezeichnete „informationelle Kapitalismus“² ist in den vergangenen drei Jahrzehnten als Antwort auf die weltweite Wirtschaftskrise Mitte der siebziger Jahre entstanden. Er umfasst – als seine zwei wichtigsten Elemente – die Globalisierung und die Informatisierung. Die Globalisierung wird heute breit diskutiert und braucht deswegen hier nicht besonders ausgeführt zu werden. Ich möchte nur einen Aspekt hervorheben, der in der Öffentlichkeit bislang zu wenig beachtet wird: Die Globalisierung hat unter anderem zu einer durchgängigen Neuordnung der Weltmärkte geführt, indem sie weltweit sehr differenzierte Märkte durchgesetzt hat, sodass auch kleinere und mittlere Unternehmen, zumal in Form von Netzwerkunternehmen, sich im globalisierten Maßstab bewegen können. Sie hat damit den Weg für die Herausbildung transnationaler Netzwerkunternehmen bereitet. Wichtiger in unserem Zusammenhang ist jedoch die Informatisierung: Sie stellt die Schlüsseltechnologien für diese neuen Märkte, für die neuartigen Unternehmen, für netzwerkförmige Organisationen und damit für die Funktionsfähigkeit einer globalisierten Wirtschaft und der sich entwickelnden Weltgesellschaft zur Verfügung.

Was heißt „Informatisierung“? Der Begriff der Informatisierung bezeichnet nicht einfach die Ausbreitung und Durchsetzung computerisierter Informations- und Kommunikationstechnologien. Informatisierungsprozesse sind vielmehr vom Beginn kapitalistischen Wirtschaftens an eine kontinuierliche Begleiterscheinung dieser Wirtschaftsform; auch Informationstechnologien gab es lange vor der Herausbildung der Computertechnologie. Neu an der Informatisierung der vergangenen drei Jahrzehnte ist dagegen, dass mit ihr globalisierte sozio-technische Systeme entstanden sind, in denen Informationen generiert, kommuniziert und in „Real Time“ verarbeitet werden. Dadurch werden Innovationen erzeugt und in einem kumulativen Rückkopplungszusammenhang wiederum für Innovationen genutzt. Daraus entstehen riesige, in ihren Dimensionen heute noch kaum absehbare, neue Produktivitätspotenziale. In etwas anderen Worten formuliert: „Informatisierung“ bedeutet die heute nahezu unbegrenzt mögliche strukturelle Verdoppelung der realen Welt in der Informationssphäre. Sachverhalte werden von vornherein als Informationsprozess verstanden, formuliert und modelliert; sie bilden in dieser Form die Ausgangsbasis für Prozesse der Reorganisation und der weiterführenden Technisierung. Der Karlsruher Philosoph Helmut Spinner formuliert das folgendermaßen: Neu sei die „technikgestützte, medienvermittelte Fähigkeit zur Wissensveränderung“; die Durchtechnisierung des Wissens in seiner Informationsform sei der Schritt von der konventionellen Technisierung zur Informatisierung.³

² Castells 2003.

³ Zitate aus Spinner 1998, S. 63 und 75. Ich habe diese Analyse der Informatisierung weiter ausgeführt in Schmiede 1996 und 2006.

Google, die Suchmaschinen, stellen eine zentrale Zugangsstelle zu dieser informatisierten Welt dar. Mit der Verlockung dieser bequemen, leicht handhabbaren Schnittstelle zwischen Mensch und Information wird deren Bedeutung für die Definition von Realität und ihrer Wahrnehmung durch uns immer größer. Soziale, ökonomische und politische Präferenzen und Entscheidungen haben sich hier zunehmend einer technischen Normierung unterzuordnen. Eine zentrale Rolle dafür spielt der von Google nie in Gänze offen gelegte Such- und Ranking-Algorithmus dieser Suchmaschine, denn er ist eine äußerst wirkmächtige Definition dessen, was Aufmerksamkeit erfordert und lohnt; bekanntlich nehmen Google-Benutzer nur die ersten 20-30 Treffer ihrer Suche noch aufmerksam zur Kenntnis.

4 INFORMATISIERUNG UND WISSEN

Mit der voranschreitenden Informatisierung verbindet sich oft die Vorstellung, das Wissen der Welt sei mehr und mehr in digitaler Form verfügbar und deshalb sei die moderne Wissensproblematik auch prinzipiell durch weitere Informatisierung lösbar. Allerdings sind trotz der greifbaren Informatisierung nicht nur die Information, sondern in zunehmendem Ausmaß auch das Wissen bedeutsamer geworden. Die Ablösung der Parole der „Informationsgesellschaft“ durch die der „Wissensgesellschaft“ signalisiert das ebenso wie die unumstritten anwachsende Rolle, die Bildung und Wissenschaft in unserer Gesellschaft spielen. Auch die Diskussionen über und die praktischen Versuche mit dem Wissensmanagement zeigen – trotz aller Ernüchterung und Unsicherheit im Umgang mit diesen Verfahren – die dem Wissen beigemessene Bedeutung.

Es lohnt sich deswegen, dem Verhältnis von Information und Wissen etwas nachzugehen. Wir schaffen Informationen, indem wir von Inhalten abstrahieren, diese Abstraktion formalisieren und positiv definieren; nur dadurch wird Information zur eindeutig bestimmten und in dieser Form technisch modellierbaren Entität. Obwohl Informationen durch diese Operation eine syntaktische Struktur erhalten, bleiben sie Rohstoff, aus dem erst Wissen geformt werden kann, aber nicht geformt werden muss. Denn Wissen bleibt immer an das Subjekt gebunden; es ist immer interpretations- und verständigungsabhängig. Wissen ist ferner an Erfahrungen gebunden, es hängt eng mit Gewissheiten und letzten Endes mit der Vernunft zusammen. Es hat eine komplexe, sich ständig in Bewegung befindende semantische Struktur. Wissen ist nicht ein Bestand, sondern es ist nur als Prozess denkbar. Es ist in praktische Zusammenhänge und Kontexte eingebunden, verändert sich deswegen auch ständig; was nicht in Handlungszusammenhängen für die wissende Person in irgendeiner Weise relevant ist, vergisst man leicht. Wissen ist daher, wie der Darmstädter Philosoph Gerhard Gamm deutlich gemacht hat, nicht positiv definierbar, sondern es ist nur in der Negation bestimmt.⁴ Der enge Zusammenhang zwischen Wissen und Person wird auch darin deutlich, dass der Status beider auf Anerkennungsprozessen beruht. Wissen ist mithin immer sozial und politisch eingebettet: „Wissen ist Macht“, lautet die Francis Bacon zugeschriebene Kernaussage.

⁴ Vgl. Gamm 2000.

Wenn Information und Wissen also in diesem Sinne ein Gegensatzpaar bilden, wie kommt es dann, dass mit der um sich greifenden Informatisierung zugleich ein sich kontinuierlich erhöhender Wissensbedarf einhergeht? Die grundlegende Antwort ist im Informatisierungsprozess selbst zu suchen: Wissensarbeit besteht im Kern in der Vermittlung zwischen abstrakter Informationswelt und der uns umgebenden, nicht informatisierten Wirklichkeit. Sie ist in wachsendem Ausmaß vonnöten, weil sie erhöhte Komplexität und zunehmende Unsicherheiten zu bearbeiten hat. Innerhalb der sich ausbreitenden Netzwerkorganisationen ist ein Mehr an medial vermittelter Kooperation und Kommunikation gefordert. Wissensarbeit ist daher in einem erheblichen Ausmaß Netzwerkarbeit; sie dient dem Ausfüllen der Lücken, die durch die Ausdünnung der Organisationen entstanden sind. Ferner muss man sich klar machen, dass durch die Informatisierung ganz neue Klassen von Wissensinhalten entstanden sind: Prozesse und Ergebnisse von Simulationen, die mathematische Beschreibung freier Formen wie zum Beispiel von Autokarosserien durch Systeme finiter Gleichungen oder das soziologische Sichtbarmachen von Sozialstrukturen durch Faktoren- und Clusteranalysen wären ohne die Nutzung von Computern gar nicht möglich.

Wissensprozesse sind daher auch nur in begrenztem Umfang technisch unterstützbar. Die Bereitstellung von „Informationen über Informationen“ (auch „Meta-Informationen“ genannt) ist noch technisch-organisatorisch lösbar; die gegenwärtigen Arbeiten am „Semantic Web“ sind wichtige Schritte in dieser Richtung. Gleichwohl geht gegenwärtig die Entwicklung stärker in die (anders begründete) Richtung der Nutzung des Web 2.0 als Ersatztechnik; Google mit seinem gigantischen Bestand an Wissen über die Kunden wird dadurch zu einem wichtigen und mächtigen Player am Wissensmarkt. Trotzdem wird damit die Wissensproblematik nicht erledigt, denn „Wissen über Wissen“ setzt mehr voraus: Erforderlich dafür ist eine Kultur der Anerkennung, der Wertschätzung, der Motivation und der Kooperation, die einen Rahmen und ein Motiv für die Generierung und die Bewahrung von Wissen bietet. Dass hier ein potenzieller Konflikt mit den Ursprüngen des Wissensmanagements besteht, springt ins Auge: Zielt dieses darauf, alle verborgenen Informationen und Wissensbestände für das Unternehmen oder andere Nutznießer nutzbar zu machen, so sind umgekehrt der ökonomische Wert und damit die Markt- und Machtposition der einzelnen Person nicht zuletzt dadurch definiert, über wieviel Information und Wissen sie als Bestandteil ihrer persönlichen Kompetenz verfügt.

5 KOMPETENZEN, CHANCEN UND GEFAHREN

Für den Umgang mit Informationen und Wissen erforderliche Fähigkeiten werden gewöhnlich mit den Begriffen „Informationskompetenz“ oder „Medienkompetenz“ beschrieben. Die damit beschriebenen Anforderungen reichen jedoch für die Bewältigung der mit Wissensarbeit verbundenen Aufgaben nicht aus, denn die für die gegenwärtige Entwicklung zentrale Kompetenz ist die an die Person gebundene Fähigkeit zur Vermittlung

zwischen Informationen und Wirklichkeit, mit anderen Worten die Kontextualisierungsfähigkeit. Diese wird jedoch durch Tendenzen zur Umwertung von Wissen gefährdet: An die Stelle von „wahr“ oder „falsch“ treten Kriterien der Zugänglichkeit wie „digital“ oder „nicht-digital“, „gegoogelt“ oder „nicht googelt“. Dagegen ist gerade das Wissen um und über das Nicht-Wissen – wie die Philosophen und Soziologen Norbert Bolz, Nina Degele und Helmut Willke betonen – von rasch wachsender Bedeutung; und dieses Wissen wird durch die genannten Zugänge überhaupt nicht eröffnet. Zentral für den menschlichen Umgang mit Wissen bleibt vielmehr die erfahrungsbasierte, Vernunft und Verstand nutzende Urteilskraft (Kant), durch die erst menschliche Autonomie und Individualität konstituiert wird. Diese Urteilskraft ist unter den heutigen Bedingungen entscheidend von der Fähigkeit abhängig, Informationen aus externen Wissensquellen zu erarbeiten, zu erschließen und zu beurteilen. Google ist wertvoll, so weit und insofern Suchmaschinen diesen Prozess unterstützen; dafür ist aber die schon eingeforderte Transparenz der Such- und Ranking-Algorithmen zentral. Die Fähigkeit zum Umgang mit nicht-digitalen Informationen hat keineswegs an Bedeutung eingebüßt, sondern bleibt für die Urteilskraft und den Umgang mit Wissen und Nicht-Wissen von zentraler Bedeutung. Dies ändert selbstverständlich nichts daran, dass Wissensprozesse durch digitale Techniken unterstützt werden können und sollten. Gegen groß angelegte Digitalisierungsprojekte, auch retrospektiver Art, ist mithin überhaupt nichts einzuwenden; sie sollten vielmehr forciert werden.

Gleichwohl sind auf der anderen Seite einige Gefahren und Fallstricke nicht zu übersehen: Die „Quatsch-Explosion“ (Gernot Wersig) droht Wissen, Erfahrung und Urteilskraft zu verdrängen.⁵ Die sachlich und sozial begründete und immer wichtiger werdende Fähigkeit zur Nicht-Information und Nicht-Kommunikation wird nach wie vor zu wenig ausgebildet. Dieser Mangel ist Bestandteil eines breiteren Defizits: Es fehlt an der Ausbildung in umfassender, die Urteilskraft stützender Informations- und Medienkompetenz, wie dies vor einigen Jahren die Dortmunder „SteFi“-Studie mit aller Deutlichkeit aufgezeigt hat.⁶ Wir hinken in der schulischen wie in der hochschulischen Ausbildung der technischen Entwicklung deutlich hinterher: Wir verweigern unseren Schülern und Studenten die Herausbildung der notwendigen Urteilsfähigkeit, indem wir die Informations- und Medienkompetenz nicht systematisch in die Schul- und Hochschulcurricula einbauen. Zwar ist die gegenwärtig in Angriff genommene technische Weiterentwicklung in der Informationserschließung erforderlich und begrüßenswert (vgl. das Web 3.0). Trotzdem bleibt die Ausbildung kritischer, auf umfassendem Wissen basierender Autonomie die Kernaufgabe, um einen angemessenen und mündigen Umgang mit den neuen Informations- und Wissenswelten zu ermöglichen.

⁵ Wersig 1987.

⁶ Sozialforschungsstelle Dortmund 2001.

6 FAZIT

Die Antwort auf die im Vortragstitel gestellte Frage kann eindeutig formuliert werden: Wir sind definitiv auf dem Weg in die „Google-Gesellschaft“! Ob diese allerdings „the information of the world at your finger-tips“ – wie das einst als Verheißung des World Wide Web formuliert wurde – verfügbar macht oder aber zur konsum-orientierten Entmündigung der Urteilskraft führen wird, ist sehr von ihrer Gestaltung abhängig. Die Ausgestaltung der Google-Gesellschaft muss die Stärkung der Autonomie der Individuen zur Leitidee haben, die ihrerseits Grundlage ist für eine anthropozentrische Technikgestaltung. Ob die „Google-Gesellschaft“ also längerfristig zum Synonym für sozialen Fortschritt oder aber zum Schimpfwort wird, hängt nicht zuletzt von unseren technischen, sozialen und politischen Bemühungen ab.

7 LITERATUR

Castells 2003

Castells, Manuel: *Das Informationszeitalter*, Bd. 3. Opladen 2003, S. 73-174.

Franck 2003

Franck, Georg: *Ökonomie der Aufmerksamkeit. Ein Entwurf*. München 2003.

Gamm 2000

Gamm, Gerhard: „Wissen und Information“. In: ders.: *Nicht nichts. Studien zu einer Semantik des Unbestimmten*, Frankfurt am Main 2000, S. 192-204.

Schmiede 1996

Schmiede, Rudi: „Informatisierung, Formalisierung und kapitalistische Produktionsweise – Entstehung der Informationstechnik und Wandel der gesellschaftlichen Arbeit“, in: ders. (Hrsg.): *Virtuelle Arbeitswelten. Arbeit, Produktion und Subjekt in der „Informationsgesellschaft“*. Berlin 1996, S. 15-47.

Schmiede 2006

Schmiede, Rudi: „Wissen und Arbeit im ‚Informational Capitalism‘“, in: Andrea Baukowitz/Thomas Berker/Andreas Boes/Sabine Pfeiffer/Rudi Schmiede/Mascha Will (Hrsg.): *Informatisierung der Arbeit – Gesellschaft im Umbruch*. Berlin 2006, S. 455-488.

Sozialforschungsstelle Dortmund 2001

Sozialforschungsstelle Dortmund: *SteFi-Studie*. URL: <http://www.stefi.de> [Stand: 20.09.2007].

Spinner 1998

Spinner, Helmut: *Die Architektur der Informationsgesellschaft. Entwurf eines wissensorientierten Gesamtkonzepts*. Bodenheim 1998.

Wersig 1987

Wersig, Gernot: „Informationsexplosion oder Quatschexplosion.“ In: *Cogito* 1 (1987), S. 44-48.

> ÜBER DIE AUTOREN

Dipl.-Medienwiss. **Markus Beiler** ist seit 2003 wissenschaftlicher Mitarbeiter am Lehrstuhl für Journalistik II des Instituts für Kommunikations- und Medienwissenschaft der Universität Leipzig. Er forscht zu den Gebieten Online-Journalismus, Suchmaschinen, internationale Mediensysteme und Media Governance. Markus Beiler hat Artikel in verschiedenen internationalen Fachzeitschriften und Sammelbänden veröffentlicht. Er ist Co-Herausgeber des Buchs *Die Macht der Suchmaschinen / The Power of Search Engines* (Köln: Herbert von Halem Verlag, 2007). Von 1999 bis 2003 hat er Medienmanagement am Institut für Journalistik und Kommunikationsforschung der Hochschule für Musik und Theater Hannover studiert. Bei der Bertelsmann Stiftung wirkte er von 2000 bis 2002 in medienpolitischen Projekten zur Ko-Regulierung des Internets mit.

Michael Buschmann ist Senior Manager Suchmaschine bei der Deutschen Telekom AG, Products & Innovation. Zu seinem Verantwortungsbereich gehören Entwicklung und Betrieb des T-Online Suchportals, Suchmaschinen-Enabling und Suchmaschinenoptimierung im T-Online-Netzwerk. 1990 begann er eine Ausbildung zum Bankkaufmann und Kredit-sachbearbeiter bei der Dresdner Bank. Dort war er bis 1998 unter anderem als Produktreferent Electronic Banking, als Produktmanager Internet und als Kaufmännischer Leiter des Projekts CyberCash im Geschäftsbereich Firmen und Institutionen, Technologie- und Innovationsmanagement (TIM) tätig. In den darauf folgenden beiden Jahren nahm er leitende Positionen bei LetsBuyIt.com Deutschland GmbH (Content Management) und bei ProSieben Digital Media GmbH (Produktion und Entwicklung) wahr und arbeitete als Berater eCommerce bei Opus5 Interaktive Medien GmbH. Bei Kirch New Media/ Kirch Intermedia war Michael Buschmann 2000 bis 2002 als Leiter Content Syndication und Mobile Services sowie als Leiter Business Development tätig. Vor Aufnahme seiner heutigen Tätigkeit war er drei Jahre bei der Deutschen Telekom als Senior Manager Produktmarketing Mobile Services unter anderem für Entwicklung und Betrieb von mobilen Services für SMS, MMS, WAP und iMode zuständig. In derselben Funktion hat er zu diversen Produktneuentwicklungen beigetragen, darunter zum Mobile eMail, wofür er als Gewinner des CONNECT Mobile Appl. Award 2005 ausgezeichnet wurde.

Dr. **Jasmin Franz** ist Director Research & Consulting bei empolis, einem Unternehmen der Bertelsmann arvato Gruppe. Dort steuert sie seit vielen Jahren in enger Zusammenarbeit mit akademischen Einrichtungen und Partnern aus der Wirtschaft nationale wie europäische Forschungsprojekte, zu denen unter anderem VICTORY (FP6: Audio-Visual ConTent search and retrieval in a distributed P2P repositORY), SEKT (Semantic Enabled Knowledge Technologies) und CIKM (Suche nach einer Methode zur Bewertung des Einflusses von Wissensmanagement auf die Innovationsfähigkeit von Unternehmen) zählen. Darüber hinaus ist Jasmin Franz aktiv in das Theseus-Forschungsprojekt eingebunden, das ebenfalls von empolis koordiniert wird. In der Position des Innovationsmanagers sorgt sie zudem für den Technologie- und Wissenstransfer sowohl empolis-intern, als auch extern durch zahlreiche Publikationen und Vorträge an Universitäten und auf nationalen und internationalen Konferenzen. Sie gehört verschiedenen Organisationskomitees zur Ausrichtung von Veranstaltungen und Konferenzen rund um das Thema Semantische Technologien an und ist auch als Autorin und Gutachterin wissenschaftlicher Beiträge diverser Magazine aktiv. Als promovierte Linguistin ist Jasmin Franz Expertin auf den Gebieten der Terminologie, Semantik, Textverständlichkeit und Informationsverarbeitung. Ihr derzeitiger Arbeitsschwerpunkt liegt im Bereich des ontologiebasierten Wissensmanagements und der angewandten Semantic Web-Technologie.

Prof. Dr. **Rainer Kuhlen** hat seit 1980 den Lehrstuhl für Informationswissenschaft an der Universität Konstanz inne. Er ist Mitglied des Fachausschusses „Kommunikation und Information“ der Deutschen UNESCO-Kommission (DUK) und hat den Deutschen UNESCO Chair in Communications (ORBICOM) inne. Darüber hinaus ist Rainer Kuhlen Mitglied im Vorstand des Hochschulverbandes für Informationswissenschaft (HI), Vorsitzender des Vereins Nethics e.V. (Informationsethik im Netz) und Sprecher des Aktionsbündnisses „Urheberrecht für Bildung und Wissenschaft“. Schließlich ist er Sachverständiger für verschiedene Bundestagsausschüsse und Enquete-Kommissionen sowie Mitglied zahlreicher Beiräte und Kommissionen für das Bundesministerium für Bildung und Forschung (BMBF), die Deutsche Forschungsgemeinschaft (DFG), die EU sowie Österreich und die Schweiz. Zu den letzten Buchpublikationen von Rainer Kuhlen gehören: *Konsequenzen von Informationsassistenten* (1999), *Informationsethik* (2004) und *Erfolgreiches Scheitern – Götterdämmerung des Urheberrechts* (in Vorbereitung, 2008). Seine Forschungs- und Lehrschwerpunkte sind Information Retrieval, Hypertext, Informationsmarkt, Informationsethik, -politik und -recht sowie kollaboratives Wissensmanagement im e-Learning.

Univ.-Prof. Dr. **Marcel Machill** ist Professor für Journalistik mit dem Schwerpunkt internationale Mediensysteme an der Universität Leipzig. Er leitet dort den Lehrstuhl für Journalistik II mit den Forschungs- und Lehrschwerpunkten (internationale) Medienpolitik und -systeme, Onlinejournalismus, Suchmaschinen, TV, Radio, Internet Governance, Medienkompetenz und journalistische Kultur(en). Vor seiner Berufung als Journalistik-Professor im April 2002 leitete Marcel Machill die Abteilung Medienpolitik der Bertelsmann Stiftung und zeichnete für das internationale Medien-Projektmanagement verantwortlich. Davor war er von 1997 bis 1999 McCloy Scholar an der Harvard University in Cambridge, USA. Marcel Machill hält akademische Grade aus drei Ländern: Bevor er nach Harvard an die John F. Kennedy School of Government kam und dort mit dem Master of Public Administration (MPA) abschloss, hat er in Paris und Dortmund Journalistik und Psychologie studiert. 1993 Diplôme an der französischen Journalistenschule Centre de Formation des Journalistes (CFJ) sowie 1994 deutsches Diplom an der Universität Dortmund. 1997 Promotion zum Dr. phil. mit einer Arbeit über die französische Medienpolitik. Seit 2003 berät Marcel Machill die Europäische Kommission bei der Evaluierung von Internet- und Internet-Governance-Projekten.

Prof. Dr. **Friedemann Mattern** ist an der ETH Zürich tätig und leitet dort das Fachgebiet „Verteilte Systeme“. Er studierte Informatik mit Nebenfach Kommunikationswissenschaften an der Universität Bonn und wurde anschließend wissenschaftlicher Mitarbeiter am Fachbereich Informatik der Universität Kaiserslautern, wo er im Sonderforschungsbereich „VLSI-Entwurf und Parallelität“ tätig war und 1989 promovierte. Zwischen 1991 und 1999 hatte er Professuren an der Universität des Saarlandes in Saarbrücken und an der Technischen Universität Darmstadt inne; an der TU Darmstadt gründete er auch das Graduiertenkolleg „Infrastruktur für den elektronischen Markt“. Nach seinem Ruf an die ETH Zürich im Jahr 1999 begann er mit dem Aufbau einer Forschungsgruppe für Ubiquitous Computing, 2002 initiiert er die Gründung des Instituts für Pervasive Computing. Friedemann Mattern ist Mitglied von acatech, korrespondierendes Mitglied der Heidelberger Akademie der Wissenschaften und gehört der Deutschen Akademie der Naturforscher Leopoldina an. Darüber hinaus ist er für diverse Institutionen als Gutachter tätig, Mitglied im Technologiebeirat mehrerer Konzerne und an verschiedenen Forschungsprojekten beteiligt. Unter anderem ist er Mitbegründer des von der ETH Zürich und der Universität St. Gallen gemeinsam getragenen M-Lab-Kompetenzzentrums. Ferner koordinierte er das von der Gottlieb Daimler- und Karl Benz-Stiftung finanzierte Ladenburger Kolleg „Leben in einer smarten Umgebung“. Sein Forschungsinteresse gilt den Gebieten „Verteilte Systeme“ und „Ubiquitous Computing“ und umfasst Modelle und Konzepte verteilter Berechnungen, Sensornetze sowie Infrastrukturmechanismen für das „Internet der Dinge“.

Prof. Dr. **Rudi Schmiede** ist seit 1987 Professor für Soziologie an der Technischen Universität Darmstadt, Fachgebiet Arbeit, Technik und Gesellschaft. Er studierte Soziologie, Sozialpsychologie, Politikwissenschaft und VWL in Frankfurt am Main, Mainz sowie an der London School of Economics. 1985 wurde er als Heisenberg-Stipendiat am Institut für Sozialforschung in Frankfurt am Main aufgenommen. Rudi Schmiede promovierte 1977 und habilitierte 1984 im Fach Soziologie. Seine Forschungs- und Publikationsschwerpunkte liegen in den Gebieten soziale Dimensionen der Informations- und Kommunikationstechnologien, geschichtliche Entwicklung von Arbeit, Arbeit in der Informationsgesellschaft, sozialstrukturelle Entwicklungstendenzen moderner Gesellschaften, Theorien der Informatisierung und des Wissens, sozial orientierte Gestaltung von IuK-Techniken, digitale Bibliotheken und integrierte wissenschaftliche Informations- und Wissenssysteme.

Prof. Dr. Dr. h.c. mult. **Wolfgang Wahlster** leitet seit 1997 als Vorsitzender der Geschäftsführung das Deutsche Forschungszentrum für Künstliche Intelligenz (DFKI GmbH) in Saarbrücken, Kaiserslautern, Bremen und Berlin. Seit 1982 ist er Inhaber des Lehrstuhls für Künstliche Intelligenz im Fachbereich Informatik der Universität des Saarlandes. Er ist Mitglied der Forschungsunion Wirtschaft-Wissenschaft der Bundesregierung und Mitglied der Königlich Schwedischen Akademie der Wissenschaften in Stockholm. Zudem ist er Vorstandsmitglied des International Computer Science Institute (ICSI) in Berkeley, Sprecher des Feldafinger Kreises und gewähltes Mitglied der Leopoldina, von acadtech und der Akademie der Wissenschaften und Literatur. Wolfgang Wahlster wurde mit seinen Innovationen als „Pionier der Computertechnik“ in die Hall of Fame des Computer-museums HNF und die Dauerausstellung des Deutschen Museums aufgenommen. Für die Neugründung von 49 Hightech-Unternehmen aus dem DFKI wurde ihm der erste deutsche Spin-Off Preis verliehen. Für seine Forschungen zu Sprachdialogsystemen wurde ihm darüber hinaus 2001 der Zukunftspreis des Bundespräsidenten und 2006 das Bundesverdienstkreuz erster Klasse verliehen. Seine aktuellen Forschungsgebiete sind multimodale Sprachdialogsysteme, das semantische Web, Fahrer- und Einkaufsassistenzsysteme, Benutzermodellierung und das Internet der Dinge.

Dr. **Stefan Wess** ist Geschäftsführer der empolis GmbH. Als Informatiker mit Nebenfach Wirtschaftswissenschaften war Stefan Wess von 1991 bis 1995 in verschiedenen internationalen IT-Projekten der Europäischen Union leitend tätig. Während dieser Zeit veröffentlichte er mehrere Bücher und zahlreiche Fachartikel. Nach seiner Promotion im Fach Informatik an der Universität Kaiserslautern wechselte er als Technical Manager Central and Southern Europe zur amerikanischen Inference Corp. in München (heute eGain Corp.). Dort übernahm er die Verantwortung für die Umsetzung von Projekten im Bereich Knowledge Management und Customer Relationship Management bei namhaften Unternehmen in ganz Europa. 1997 wurde er Geschäftsführer der tec:inno GmbH (seit 2000 vollständig in die empolis GmbH integriert), wo er die strategische Neuausrichtung des IT-Unternehmens in den Bereich Knowledge Management umsetzte. Stefan Wess ist Mitglied im Aufsichtsrat des Deutschen Forschungszentrums für Künstliche Intelligenz in Kaiserslautern, Mitglied im Kuratorium des Fraunhofer-Instituts für Intelligente Analyse- und Informationssysteme IAIS in Sankt Augustin und Mitglied im Arbeitskreis Knowledge Management der BITKOM.

Martin Zenker, M.A., arbeitet seit 2004 am Lehrstuhl für Journalistik II des Institutes für Kommunikations- und Medienwissenschaft der Universität Leipzig bei Prof. Dr. Marcel Machill, aktuell als wissenschaftlicher Mitarbeiter im Forschungsprojekt „Journalistische Recherche im Internet“. Von 1997 bis 2003 studierte er Kommunikations- und Medienwissenschaft sowie Amerikanistik und Soziologie an der Universität Leipzig. Seit 1996 ist er als Freier Mitarbeiter im lokalen Printjournalismus tätig. Fotografische Tätigkeit u. a. für das Amt für Geoinformationswesen der Bundeswehr (Afghanistan, 2003) sowie diverse Printmedien. 2006 Lehrauftrag für „Bildreportage Online“ an der Universität Leipzig.

> acatech – DEUTSCHE AKADEMIE DER TECHNIKWISSENSCHAFTEN

acatech vertritt die Interessen der deutschen Technikwissenschaften im In- und Ausland in selbstbestimmter, unabhängiger und gemeinwohlorientierter Weise. Als Arbeitsakademie berät acatech Politik und Gesellschaft in technikwissenschaftlichen und technologiepolitischen Zukunftsfragen. Darüber hinaus hat es sich acatech zum Ziel gesetzt, den Wissenstransfer zwischen Wissenschaft und Wirtschaft zu erleichtern und den technikwissenschaftlichen Nachwuchs zu fördern. Zu den Mitgliedern der Akademie zählen herausragende Wissenschaftler aus Hochschulen, Forschungseinrichtungen und Unternehmen. acatech finanziert sich durch eine institutionelle Förderung von Bund und Ländern sowie durch Spenden und projektbezogene Drittmittel. Um die Akzeptanz des technischen Fortschritts in Deutschland zu fördern und das Potenzial zukunftsweisender Technologien für Wirtschaft und Gesellschaft deutlich zu machen, veranstaltet acatech Symposien, Foren, Podiumsdiskussionen und Workshops. Mit Studien, Empfehlungen und Stellungnahmen wendet sich acatech an die Öffentlichkeit. acatech besteht aus drei Organen: Die Mitglieder der Akademie sind in der Mitgliederversammlung organisiert; der Vorstand, der von den Mitgliedern der Akademie bestimmt wird, lenkt die Arbeit; ein Senat mit namhaften Persönlichkeiten aus Industrie, Wissenschaft und Politik berät acatech in Fragen der strategischen Ausrichtung und sorgt für den Austausch mit der Wirtschaft und anderen Wissenschaftsorganisationen in Deutschland. Die Geschäftsstelle von acatech befindet sich in München; zudem ist acatech mit einem Hauptstadtbüro in Berlin vertreten.

Weitere Informationen unter www.acatech.de

> acatech diskutiert

Die Reihe „acatech diskutiert“ dient der Dokumentation von Symposien, Workshops und weiteren Veranstaltungen der Deutschen Akademie der Technikwissenschaften. Darüber hinaus werden in der Reihe auch Ergebnisse aus Projektarbeiten bei acatech veröffentlicht. Die Bände dieser Reihe liegen generell in der inhaltlichen Verantwortung der jeweiligen Herausgeber und Autoren.