

Ilmenauer Beiträge zur Wirtschaftsinformatik

Herausgegeben von U. Bankhofer, V. Nissen

D. Stelzer und S. Straßburger

Bastian Eine, Bernd Markscheffel

**Alternative Websuchdienste:
Überblick und Vergleich**

Arbeitsbericht Nr. 2011-01, Januar 2011



Technische Universität Ilmenau
Fakultät für Wirtschaftswissenschaften
Institut für Wirtschaftsinformatik

Autor: Bastian Eine, Bernd Markscheffel

Titel: Alternative Websuchdienste: Übersicht und Vergleich

Ilmenauer Beiträge zur Wirtschaftsinformatik Nr. 2011-01, Technische Universität Ilmenau, 2011

ISSN 1861-9223

ISBN 978-3-938940-32-7

urn:nbn:de:gbv:ilm1-2011200032

© 2011 Institut für Wirtschaftsinformatik, TU Ilmenau

Anschrift Technische Universität Ilmenau, Fakultät für Wirtschaftswissenschaften,
Institut für Wirtschaftsinformatik, PF 100565, D-98684 Ilmenau.
<http://www.tu-ilmenau.de/wid/forschung/ilmenauer-beitraege-zur-wirtschaftsinformatik>

Gliederung

| | |
|---|----|
| Gliederung..... | ii |
| Abbildungsverzeichnis..... | iv |
| Tabellenverzeichnis..... | iv |
| Abkürzungsverzeichnis..... | v |
| 1 Einleitung..... | 1 |
| 1.1 Problemstellung..... | 1 |
| 1.2 Zielsetzung..... | 3 |
| 1.3 Methodik..... | 3 |
| 1.4 Aufbau der Arbeit..... | 4 |
| 2 Grundlagen..... | 6 |
| 2.1 Websuchdienste..... | 6 |
| 2.2 Besonderheiten des Web Information Retrieval..... | 8 |
| 2.2.1 Struktur des World Wide Web..... | 8 |
| 2.2.2 Daten des World Wide Web..... | 10 |
| 2.2.3 Nutzer von Websuchdiensten..... | 13 |
| 2.2.4 Schlussfolgerungen..... | 13 |
| 2.3 Markt für Websuchdienste..... | 13 |
| 2.3.1 Verteilung der Anteile..... | 14 |
| 2.3.2 Entwicklung..... | 15 |
| 2.3.3 Chancen für alternative Websuchdienste..... | 16 |
| 2.3.4 Schlussfolgerungen..... | 17 |
| 2.4 Kennzahlen für einen Vergleich von Websuchdiensten..... | 18 |
| 2.4.1 Linktopologische Kennzahlen..... | 20 |
| 2.4.1.1 Web Impact Factor..... | 21 |

| | |
|--|----|
| 2.4.1.2 PageRank..... | 22 |
| 2.4.2 Nutzungskennzahlen..... | 22 |
| 2.4.3 Schlussfolgerungen..... | 25 |
| 3 Überblick über Klassifizierungsansätze und Kategorien..... | 25 |
| 3.1 Klassifizierungsansätze..... | 25 |
| 3.1.1 Nutzerführung..... | 25 |
| 3.1.2 Allgemeine und spezielle Suche..... | 26 |
| 3.1.3 Manuelle und automatische Inhaltserschließung..... | 27 |
| 3.1.4 Deep Web-Suche..... | 27 |
| 3.1.5 Soziale Suche..... | 27 |
| 3.1.6 Echtzeitsuche..... | 29 |
| 3.1.7 Semantische Suche..... | 29 |
| 3.1.8 Visuelle Suche..... | 30 |
| 3.1.9 Personalisierte Suche..... | 31 |
| 3.1.10 Lokale Suche..... | 31 |
| 3.1.11 Schlussfolgerungen..... | 32 |
| 3.2 Kategorien..... | 32 |
| 3.2.1 Bilder-, Video- und Audiosuchdienste..... | 33 |
| 3.2.2 Personensuchdienste..... | 34 |
| 3.2.3 Frage-Antwort-Dienste..... | 34 |
| 3.2.4 Social Bookmarking-Dienste..... | 35 |
| 3.2.5 Blog- und Twittersuchdienste..... | 35 |
| 3.2.6 Nachrichtensuchdienste..... | 36 |
| 3.2.7 Wissenschaftssuchdienste..... | 37 |
| 3.2.8 Schlussfolgerungen..... | 37 |
| 4 Vergleich alternativer Websuchdienste..... | 38 |

| | |
|---|----|
| 4.1 Ermittlung der Grundgesamtheit..... | 38 |
| 4.1.1 Websuchdienstlisten..... | 41 |
| 4.1.2 Suchanfragen an Google..... | 44 |
| 4.2 Kennzahlenerhebung..... | 47 |
| 4.2.1 Web Impact Factor..... | 48 |
| 4.2.2 PageRank..... | 49 |
| 4.2.3 Alexa Traffic Rank..... | 49 |
| 4.3 Berechnung der Gesamtwerte und Erstellung der Rankings..... | 49 |
| 4.4 Interpretation der Rankings..... | 61 |
| 5 Zusammenfassung und Ausblick..... | 65 |
| Literaturverzeichnis..... | 67 |

Abbildungsverzeichnis

| | |
|--|----|
| Abb. 1: Bow-Tie-Struktur..... | 9 |
| Abb. 2: Nutzung von Websuchdiensten auf dem deutschen Markt..... | 14 |
| Abb. 3: Nutzung von Websuchdiensten auf dem U.S. Markt..... | 14 |
| Abb. 4: Entwicklung der Marktanteile auf dem deutschen Markt..... | 15 |
| Abb. 5: Ausschnitt aus der Websuchdienstliste von Pandia Powersearch..... | 41 |
| Abb. 6: Ausschnitt aus dem Suchergebnis zur Kategorie Bildersuchdienste..... | 46 |

Tabellenverzeichnis

| | |
|--|----|
| Tab. 1: Ranking der Bildersuchdienste..... | 51 |
| Tab. 2: Ranking der Videosuchdienste..... | 52 |

| | |
|---|----|
| Tab. 3: Ranking der Audiosuchdienste..... | 53 |
| Tab. 4: Ranking der Personensuchdienste..... | 54 |
| Tab. 5: Ranking der Frage-Antwort-Dienste..... | 55 |
| Tab. 6: Ranking der Social Bookmarking-Dienste..... | 56 |
| Tab. 7: Ranking der Blogsuchdienste..... | 57 |
| Tab. 8: Ranking der Twittersuchdienste..... | 58 |
| Tab. 9: Ranking der Nachrichtensuchdienste..... | 59 |
| Tab. 10: Ranking der Wissenschaftssuchdienste..... | 60 |

Abkürzungsverzeichnis

| | |
|------|----------------------------|
| HTML | Hypertext Markup Language |
| IF | Impact Factor |
| IR | Information Retrieval |
| SEM | Search Engine Marketing |
| SEO | Search Engine Optimization |
| URL | Uniform Resource Locator |
| WIF | Web Impact Factor |
| WWW | World Wide Web |

Zusammenfassung: Websuchdienste sind durch die unüberschaubare Menge an Informationen im World Wide Web (WWW) zu einem unverzichtbaren Werkzeug bei der Informationssuche geworden. Eine Reihe von alternativen Websuchdiensten versuchen dem Nutzer besonders bei der themen- oder formatspezifischen Suche bessere Ergebnisse als die der allgemeinen Suchmaschinen zu liefern. In dieser Arbeit wird ein Überblick über Websuchdienste gegeben, die zum derzeitigen Stand neben den marktbeherrschenden allgemeinen Suchmaschinen und den allgemeinen Verzeichnisdiensten im WWW verfügbar sind. Dazu wurden alternative Websuchdienste innerhalb ausgewählter Kategorien miteinander verglichen und die populärsten Websuchdienste der jeweiligen Kategorien ermittelt. Für den Vergleich der alternativen Websuchdienste wurden Kennzahlen aus der Webometrie herangezogen, die zur Analyse der Strukturen, Nutzung und Inhalte des WWW entwickelt worden sind.

Schlüsselworte: Alternative Websuchdienste, Webometrie, Web Impact Factor, Google PageRank, Alexa Traffic Rank

1 Einleitung

1.1 Problemstellung

Die Menge an Informationen im WWW ist schwer abschätzbar [Lewa05, 41] und für den allgemeinen Nutzer unüberschaubar. Eine Vielzahl von Websuchdiensten¹ werden angeboten, um den Nutzer bei der Informationssuche im WWW zu unterstützen.² Die drei Suchmaschinen³ Google, Bing (Microsoft) und Yahoo besitzen für den deutschen Markt einen Marktanteil von über 90%⁴ und werden am häufigsten für die Recherche im WWW genutzt. Neben den populärsten Suchdiensten, den allgemeinen⁵ Suchmaschinen und den allgemeinen Verzeichnisdiensten (Kataloge)⁶, kann auf weitere, alternative Websuchdienste⁷ zurückgegriffen werden. Alternative Websuchdienste verwenden verschiedene Suchansätze und -techniken, die sich von denen der allgemeinen Suchmaschinen und den allgemeinen Verzeichnisdiensten unterscheiden.

Bei dem Versuch Websuchdienste zu systematisieren, ergibt sich das Problem, dass eine Einordnung von Websuchdiensten in übersichtliche Kategorien durch die Vielfalt von unterschiedlichen Suchansätzen und der ständigen Weiterentwicklung der Websuchdienste erschwert wird. Es lassen sich jedoch einige Merkmale zur Abgrenzung von Websuchdiensten bestimmen. So kann z. B. zwischen manueller und automatischer

¹ Unter Websuchdienste werden in dieser Arbeit Dienste im WWW verstanden, die den Nutzer bei der Informationssuche im WWW unterstützen. Der Begriff Suchdienste wird in dieser Arbeit synonym verwendet und bezieht sich ebenfalls auf das WWW.

² Siehe z. B. <http://www.searchenginecolossus.com/>, <http://www.thesearchenginelist.com/>, <http://www.suchlexikon.de/> oder <http://www.klug-suchen.de>, Abruf: 2010-08-05

³ „Unter Suchmaschinen wird hier die algorithmische Suchlösung verstanden, das heißt, Dokumente werden automatisch mittels Crawlern erfasst, automatisch erschlossen und auf eine Nutzeranfrage hin werden passende Dokumente – nach der angenommenen Relevanz sortiert – zurückgegeben.“ [Lewa08, 1]

⁴ Vgl. Web-Barometer unter <http://www.webhits.de>; für den U.S. Markt siehe Nielsen MegaView Search unter http://en-us.nielsen.com/content/nielsen/en_us/insights/rankings/internet.html, Abruf: 2010-10-17

⁵ Als allgemein werden diejenigen Suchdienste bezeichnet, die versuchen alle Dokumente im WWW zu erfassen; mit der Ausnahme von unerwünschten Inhalten, wie z. B. Spam [Lewa09c, 54].

⁶ „In Katalogen werden Websites manuell erfasst, mit einer kurzen Beschreibung versehen und in ein Klassifikationssystem einsortiert.“ [Lewa08, 1]

⁷ Unter alternativen Websuchdiensten werden in dieser Arbeit Suchdienste verstanden, die sich in ihrem Suchkonzept von den allgemeinen Suchmaschinen und den allgemeinen Verzeichnisdiensten unterscheiden.

Inhaltserschließung unterschieden werden [Lewa05, 24 ff.]. Ebenso kann eine Einteilung nach universellem und spezialisiertem Anwendungskontext durchgeführt werden [GBRi09, 49].

In Lewandowskis Ausführungen zu Spezialsuchmaschinen⁸ werden einige Nachteile von allgemeinen Suchmaschinen beschrieben und die Vorteile von alternativen Websuchdiensten verdeutlicht. Obwohl die allgemeinen Suchmaschinen durch algorithmische Verfahren versuchen, alle Dokumente im WWW zu erfassen, erreichen diese keine vollständige Abdeckung des WWW [Lewa05, 41 ff.]. Eine Vielzahl von Informationen befindet sich z. B. in zugangsbeschränkten Datenbanken oder ist auf Grund weiterer technischer Hindernisse durch allgemeine Suchmaschinen nicht erfassbar.⁹ Zudem sind die allgemeinen Suchmaschinen an den Bedürfnissen der Mehrheit der Nutzer angepasst, so dass es z. B. exakte Nutzeranfragen erfordert, um zu spezifischen Treffern zu gelangen [Lewa09c, 55]. Vorteile bei der Informationssuche mit alternativen Websuchdiensten können dann entstehen, wenn sich diese Websuchdienste auf bestimmte Bereiche des WWW spezialisieren. Die Spezialisierung kann sich z. B. auf bestimmte Themen, auf bestimmte Dokumenttypen und -formate, auf zeitkritische¹⁰ Informationen oder auf die Verbindung von Inhalten aus dem freien Web mit Inhalten aus dem Deep Web beziehen [Lewa09c, 57 ff.]. An diese Spezialisierungen können alternative Websuchdienste ihre Verfahren zur Inhaltserschließung, ihre bereitgestellten Suchfunktionen und ihre Darstellung der Suchergebnisse anpassen. Beispiele für alternative Websuchdienste sind Multimediasuchdienste¹¹, Personensuchdienste,¹² Blogsuchdienste¹³, Nachrichtensuchdienste¹⁴ und Wissenschaftssuchdienste¹⁵. Weitere, sich aktuell entwickelnde

⁸ „Spezialsuchmaschinen (...) sind solche, die sich thematisch oder anhand formaler Dokumentenmerkmale (Bsp. Dateityp) beschränken.“ [Lewa09c, 57]

⁹ Dieser Teil des WWW wird als Deep Web bezeichnet [Lewa09c, 54].

¹⁰ Zeitkritisch bedeutet, dass diese Art von Informationen nach kurzer Zeit an Aktualität verlieren.

¹¹ Multimediasuchdienste durchsuchen das WWW nach Bildern, Videos und Audios [Lewa09c, 65 ff.].

¹² Personensuchdienste durchsuchen das WWW nach Informationen über Personen [Karz10, o. S.].

¹³ Blogsuchdienste durchsuchen das WWW nach Blogs, Blogbeiträgen oder Blogautoren [Lewa09c, 60].

¹⁴ Nachrichtensuchdienste durchsuchen ausgewählte Quellen des WWW nach aktuellen Meldungen [Lewa09c, 59].

¹⁵ Wissenschaftssuchdienste durchsuchen das WWW nach wissenschaftlichen Dokumenten [Lewa09c, 62 f.].

Suchansätze sind z. B. die Soziale Suche¹⁶, Echtzeitsuche¹⁷, Personalisierte Suche¹⁸, Semantische Suche¹⁹ und die Visuelle Suche²⁰.

Die Vorteile alternativer Websuchdienste gegenüber den allgemeinen Suchmaschinen und den allgemeinen Verzeichnisdiensten verdeutlichen, dass eine Betrachtung von Websuchdiensten über die marktbeherrschenden allgemeinen Websuchdienste hinaus gehen sollte. Auf Grund der Vielzahl verschiedener Suchansätze und -techniken von alternativen Websuchdiensten, die auch zunehmend in das Angebot der großen Suchmaschinenanbieter integriert werden²¹, ist es von Interesse, sich einen Überblick über alternative Websuchdienste zu verschaffen und deren Konzepte hervorzuheben.

1.2 Zielsetzung

Ziel dieser Arbeit ist es, einen Überblick über alternative Websuchdienste zum derzeitigen Stand zu geben. Dazu werden Rankings in ausgewählten Websuchdienstkategorien mit der Hilfe von webometrischen Kennzahlen erstellt.

1.3 Methodik

Ein Überblick über Klassifizierungsansätze und Kategorien von Websuchdiensten wird durch Auswertung der Fachliteratur gegeben. Die Auswahl der Websuchdienstkategorien orientiert sich an den von Lewandowski [Lewa09c, 59 ff.], Griesbaum, Bekavac und Rittberger [GBRi09, 18 ff.; Beka09, o. S.] vorgestellten Websuchdienstkategorien.

Für die Erstellung der Rankings in den ausgewählten Kategorien wird eine Grundgesamtheit an alternativen Websuchdiensten ermittelt. Dazu werden zum einen

¹⁶ In sozialen Suchansätzen werden die Suchanfragen dadurch beantwortet, dass andere Nutzer, die zumeist ähnliche Interessen teilen, sich bei diesen Diensten mit einbringen [MSHP09, 6].

¹⁷ Mit der Echtzeitsuche werden Inhalte im WWW möglichst zeitnah nach deren Erstellung erfasst.

¹⁸ Ansätze zur Personalisierten Suche bieten dem Nutzer Möglichkeiten, die Suche durch Anpassungen an das eigene Suchverhalten zu verbessern [Grie07, 13].

¹⁹ Für semantische Suchansätze wird versucht, den Inhalten im WWW und deren Beziehungen untereinander durch zusätzliche Beschreibungen eine Bedeutung zu geben, die auch von automatischen Verfahren erkannt und verarbeitet werden kann [BHLa01, o. S.].

²⁰ Bei visuellen Suchansätzen wird die Suche nach Informationen durch graphische Darstellungen ergänzt [WBHÖ+09, 252].

²¹ Beispiele für solche Integrationen sind die Google Blogsuche unter <http://blogsearch.google.de/>, die Yahoo Nachrichtensuche unter <http://de.news.search.yahoo.com/> oder die Bing Bildersuche unter <http://www.bing.com/images>, Abruf: 2009-12-15

fachbezogene Quellen im WWW herangezogen, die eine Liste alternativer Websuchdienste zusammengestellt haben, und zum anderen Suchergebnisse ausgewertet, die mit Hilfe von Suchanfragen an die allgemeine Suchmaschine Google abgerufen werden. Die Auswertung der Websuchdienstlisten und der Suchergebnisse wird nach Kriterien durchgeführt, die sich zum einen aus den Ausführungen zu den einzelnen Websuchdienstkategorien ergeben und zum anderen im Rahmen dieser Arbeit aufgestellt werden.

Um die ermittelte Grundgesamtheit an alternativen Websuchdiensten in eine bewertete Reihenfolge innerhalb der Kategorien zu bringen, wird die Popularität der Websuchdienste mit Hilfe webometrischer²² Kennzahlen verglichen. Als webometrische Kennzahlen dienen in dieser Arbeit zum einen die linktopologischen Kennzahlen Web Impact Factor (WIF)²³ und PageRank²⁴ und zum anderen die Nutzungskennzahl Alexa Traffic Rank²⁵. Für jeden Websuchdienst der Grundgesamtheit werden die Werte dieser ausgewählten Kennzahlen abgerufen und anschließend normiert, um daraus einen Gesamtwert für jeden Websuchdienst zu berechnen. Anhand dieser Gesamtwerte wird innerhalb jeder Websuchdienstkategorie ein Ranking der Websuchdienste aufgestellt. Die ersten zehn Websuchdienste jeder Kategorie werden in einer tabellarischen Übersicht dargestellt.

1.4 Aufbau der Arbeit

Im Anschluss an die Einleitung wird im zweiten Kapitel auf den Begriff Websuchdienste eingegangen und das wissenschaftliche Umfeld erläutert. Ebenso werden in Kapitel zwei die Besonderheiten des Web Information Retrieval, der Markt für Websuchdienste und die Kennzahlen für einen Vergleich von Websuchdiensten dargestellt. In Kapitel drei wird ein Überblick über Klassifizierungsansätze und Kategorien alternativer Websuchdienste gegeben. Kapitel vier beschreibt die Ermittlung der Grundgesamtheit an alternativen Websuchdiensten, deren Vergleich anhand von webometrischen Kennzahlen und die Interpretation der Rankings. Kapitel fünf dient der Zusammenfassung der erarbeiteten

²² Mit webometrischen Kennzahlen werden Struktur, Nutzung und Inhalte des WWW untersucht [Björ04; The109].

²³ Der WIF misst die Attraktivität einer Website, indem das Verhältnis von eingehenden Verlinkungen zu der Anzahl der Webseiten, aus denen eine Website besteht, betrachtet wird [Ingw98, 236 ff.].

²⁴ Der PageRank wertet ebenfalls Linkstrukturen von Websites aus. Zusätzlich zu der Anzahl der Verlinkungen wird die Gewichtung der Verlinkungen berücksichtigt [PBMW98, 3].

²⁵ Für den Alexa Traffic Rank wird die Anzahl von Besuchern einer Website und die Anzahl von Seitenaufrufen innerhalb einer Website ausgewertet. Vgl. http://www.alexa.com/help/traffic_learn_more, Abruf: 2010-09-21

Ergebnisse, der kritischen Reflexion der Arbeit und gibt einen Ausblick auf zukünftige Entwicklungen und offene Forschungsthemen.

2 Grundlagen

2.1 Websuchdienste

Bevor auf alternative Websuchdienste eingegangen werden kann, muss zuerst der Begriff Websuchdienste näher bestimmt und in ein wissenschaftliches Umfeld eingeordnet werden.

Der Begriff Websuchdienste ist in der Literatur nur selten zu finden²⁶ und besitzt keine allgemein anerkannte Definition. Websuchdienste kann als Oberbegriff für die unterschiedlichen Suchansätze und -kategorien verstanden werden, die in den Abschnitten 3.1 und 3.2 erläutert werden. Durch den vorangestellten Zusatz Web wird der Bezug auf die im WWW bereitgestellten Suchdienste verdeutlicht. In der Literatur sind unterschiedliche Synonyme (z. B. Websuchwerkzeug, Websuchhilfe, Websuchlösung), ebenso in englischer Sprache (z. B. web search tools, web search, web search service), für den Begriff Websuchdienste zu finden. Zudem wird häufig nicht zwischen den verschiedenen Ansätzen und Kategorien von Websuchdiensten unterschieden, sondern repräsentativ der Begriff Suchmaschine verwendet.

Websuchdienste werden von verschiedenen Bereichen der Wissenschaft erforscht. Da Informationen weitestgehend maschinell verarbeitet werden, sind Websuchdienste eng mit der Informatik verbunden.²⁷ Allerdings sind Websuchdienste als Hilfsmittel für den Menschen entworfen worden und stehen in Interaktion mit diesem. Deshalb sind „neben den Kernbereichen Informatik und Informationswissenschaft (...) vor allem die Geistes- und Gesellschaftswissenschaften zu nennen“ [Lewa08, 7]. Während die Informatik den technischen Bereich der Websuchdienste forciert, werden im Kontext der Informationswissenschaft die Bedürfnisse des Nutzers in die Betrachtung mit einbezogen [Lewa05, 15]. In den Geistes- und Gesellschaftswissenschaften stehen ebenfalls die Nutzer im Vordergrund. Neben den individuellen Nutzerbedürfnissen werden hierbei vor allem die Bedeutung und der Einfluss der Websuchdienste auf Gesellschaft und Wirtschaft untersucht. Schließlich können auch Rechtswissenschaften ein Teil der Websuchdienstforschung sein. Dabei kann es sich z. B. um datenschutzrechtliche Fragen handeln, wenn

²⁶ Der Begriff Websuchdienste ist in Beiträgen von Griesbaum und Bekavac zu finden, wird dort aber nicht definiert [GBRi09; GrBe04a; GrBe04b].

²⁷ Vgl. zum folgenden Abschnitt [SaMc87, Vorwort].

bei der Nutzung von Websuchdiensten personenbezogene Daten anfallen und diese von den Websuchdiensteanbietern ausgewertet werden [Weic09, 1]. Lewandowski empfiehlt eine interdisziplinäre Ausrichtung der Forschung zu Websuchdiensten, da einzelne Wissenschaftsfelder nicht das gesamte Themenfeld von Websuchdiensten abdecken können [Lewa08, 7].

Die Entwicklung der Websuchdienste ist eng mit der des WWW verknüpft. Nach der Vorstellung des Konzepts durch Tim Berners-Lee im Jahr 1989 [Bern89, o. S.] wuchs das WWW auf eine unüberschaubare Anzahl von verknüpften Dokumenten an. Deshalb erforderte diese Situation ein Werkzeug, welches diese Dokumente durchsuchbar und enthaltene Informationen sowie die Dokumente selbst leichter auffindbar machen sollte. Die Anwendung solcher Hilfsmittel war bereits abseits des WWW aus dem Einsatz in anderen großen Dokumentensammlungen, wie z. B. in spezialisierten Datenbanken, bekannt.

Das Information Retrieval (IR) beschäftigt sich mit der Forschung zu den Verfahren zum Wiederauffinden von Informationen und wurde als Begriff erstmals von Calvin Mooers im Jahr 1950 eingeführt [Stoc07, 38]. Anfänglich wurde das „(Wieder)auffinden“ von wissenschaftlicher Literatur untersucht [Ferb03, 29]. Eine Definition des IR durch Salton und McGill zeigt, welche Teilbereiche beim Umgang mit Informationen berücksichtigt werden müssen: „Gegenstand des IR ist die Repräsentation, Speicherung und Organisation von Informationen und der Zugriff auf Informationen“ [SaMc87, 1].

In Bezug auf Websuchdienste hat sich mit dem Web IR ein gesonderter Forschungsbereich entwickelt. Das Web IR knüpft an das klassische IR an und hat sich durch die Adaption metrischer Konzepte auf das WWW herausgebildet [Lewa05, 13 f.]. Struktur, Daten und Nutzung des WWW unterscheiden sich jedoch von denen der Datenbanken, mit welchen sich das klassische IR beschäftigt. In einer Definition von Manning, Raghavan und Schütze aus dem Jahr 2008 zum IR sind bereits Hinweise auf strukturelle Besonderheiten des WWW erkennbar: „IR is finding material (usually documents) of an unstructured nature (usually text) that satisfies an information need from within large collections (usually stored in computers)“ [MRSc08, 1]. Hier werden Unstrukturiertheit der Inhalte und Größe der Dokumentensammlung genannt. Im folgenden Abschnitt 2.2 wird auf weitere bei der Websuche zu berücksichtigende Eigenschaften des WWW und dessen Nutzer eingegangen, so dass die Unterschiede zwischen dem klassischen IR und dem Web IR erkennbar werden.

2.2 Besonderheiten des Web Information Retrieval

Das Web IR kann auf Wissen und Verfahren aus dem klassischen IR zurückgreifen. Eine Übertragung dieser Bestände auf Websuchdienste sollte allerdings unter Beachtung der folgenden Besonderheiten des Web IR erfolgen. In diesem Abschnitt wird dargestellt, welche Probleme durch die Struktur und den Daten des WWW sowie den Nutzern von Websuchdiensten entstehen können.

Die Besonderheiten des Web IR können sich auf unterschiedliche Bereiche beziehen. Häufig wird in der Literatur zwischen Eigenschaften bezüglich der Daten selbst und den Eigenschaften, die den Nutzer und seiner Interaktion mit dem WWW betreffen, unterschieden.²⁸ Eine weitere systematische Unterscheidung nennt Lewandowski.²⁹ Nach Lewandowski lassen sich Unterschiede zwischen dem klassischen IR und dem Web IR in vier Klassen einteilen: Unterschiede bezüglich des Dokumentenkorpus, der Inhalte, der Nutzer und der Eigenarten des IR Systems selbst. Die folgende Darstellung der Besonderheiten des Web IR ist in Struktur des WWW, Daten des WWW und Nutzer von Websuchdiensten unterteilt.

2.2.1 Struktur des World Wide Web

Das WWW besteht aus einer Vielzahl von Dokumenten, die dezentral gespeichert sind [BaRi99, 368]. Die Verknüpfung der Dokumente untereinander ist unterschiedlich stark ausgeprägt und kann mit der Bow-Tie-Struktur (siehe Abb. 1) erklärt werden.³⁰

²⁸ Vgl. z. B. [CrHa09, 85] oder [BaRi99, 368].

²⁹ Vgl. zum folgenden Abschnitt [Lewa05, 71 ff.].

³⁰ Vgl. zum folgenden Abschnitt [BKMR+00, o. S.; Lewa05, 45 f.].

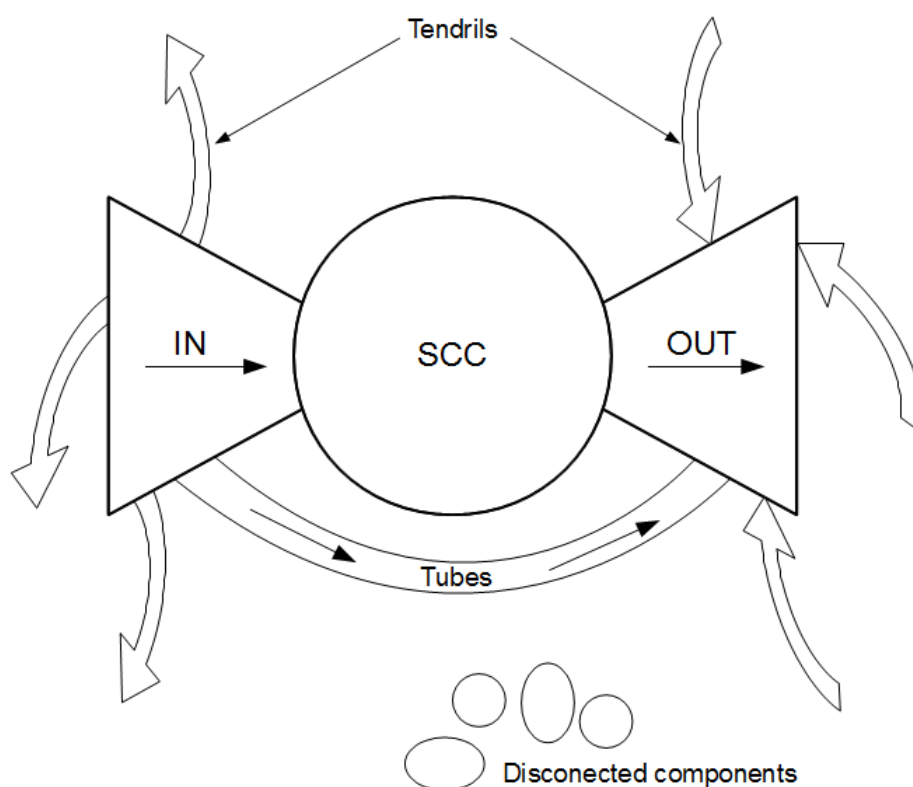


Abb. 1: Bow-Tie-Struktur³¹

Hierbei wird von einem stark untereinander verbundenen Kernbestand an Dokumenten ausgegangen (Strongly Connected Core - SCC). Neben diesem Kernbereich gibt es einen Bereich von Dokumenten, der auf den Kernbereich verweist (IN), und einen weiteren Bereich von Dokumenten, auf den der Kernbereich selbst verweist (OUT). Vereinzelt existieren Dokumente, die entweder nur eine einseitige Verbindung mit einem In- oder Out-Bereich haben (Tendrils) oder eine Verbindung zwischen dem In- und Out-Bereichen herstellen (Tubes). Schließlich existieren auch Dokumente, die gar keine Verbindung zu einem der drei Bereiche aufweisen und damit als unverbundene Dokumente bezeichnet werden können (Disconnected components).

Für Websuchdienste kann dies bei der Erschließung von Dokumenten zu Problemen führen. Schwach verknüpfte oder isolierte Dokumente werden durch automatische Verfahren nur mit geringer Wahrscheinlichkeit bzw. überhaupt nicht erschlossen, da automatische Verfahren Dokumente über Verknüpfungen zwischen den Dokumenten finden. Eine manuelle Erschließung setzt die Kenntnis der Adresse (Uniform Resource

³¹ Vgl. [BKMR+00, o. S.]

Locator - URL) des Dokuments voraus und kann auf Grund der großen Anzahl von Dokumenten im WWW nur für einen Teil der Dokumente durchgeführt werden.

Ebenso existiert im WWW eine Vielzahl von Dokumenten, die einer Zugangsbeschränkung (z. B. Kosten- oder Registrierungspflicht) unterliegen oder erst durch die Formulierung einer Anfrage abgerufen werden können (z. B. Dokumente in Online-Datenbanken). Diese Dokumente sind für Websuchdienste durch einen erschwerten Zugriff problematisch.

Auf Grund der Struktur des WWW ist also zum einen die Anzahl der Dokumente im WWW nicht bestimmbar und zum anderen können Websuchdienste keine vollständige Abdeckung des WWW gewährleisten.

2.2.2 Daten des World Wide Web

Das WWW besteht aus Dokumenten mit sehr heterogenen Eigenschaften. Neben den heterogenen Eigenschaften der Dokumente können auch gleiche und ähnliche Inhalte [Lewa05, 72 f.] sowie Spam [HMSi02, 1] zu Problemen bei Websuchdiensten führen. Diese Besonderheiten sind auf die großen Freiheiten für Nutzer und Anbieter bei der Bereitstellung von Informationen im WWW zurückzuführen.³² Die Freiheiten resultieren aus dem freien Zugang zum WWW und dem Fehlen von inhaltlichen und formalen Anforderungskriterien, einer Qualitätskontrolle oder einer redaktionellen Auswahl. Die Heterogenität bezieht sich auf die Dokumenttypen und -formate, die Strukturierung innerhalb der Dokumente, die verwendeten Sprachen sowie die Qualität und Aktualität der Dokumenteninhalte [BaRi99, 368; Lewa05, 72].

Dokumente mit unterschiedlichen Dokumenttypen und -formaten sowie deren Inhalte müssen durch Verfahren erfasst werden, die sich von gewöhnlichen Verfahren zur Erschließung von Textinhalten unterscheiden. Dokumente mit Multimediainformationen (Bilder, Videos, Audios) werden z. B. mit Hilfe von Metainformationen erschlossen [Lewa05, 71]. Diese Zusatzinformationen werden aus Dateinamen, Kommentaren, Verlinkungen, Ankertexten oder weiteren Hinweisen entnommen [Ferb03, 286]. Ebenso können die Dokumente selbst aus Dateien mit unterschiedlichen Formaten zusammengesetzt (Granularität) oder zur besseren Lesbarkeit in mehrere Teile zerlegt sein.³³ Dadurch

³² Vgl. zum folgenden Abschnitt [BaRi99, 7].

³³ Vgl. zum folgenden Abschnitt [Ferb03, 288].

kommt es vor, dass der Zusammenhang zwischen einzelnen Teilen des Dokuments nicht mehr von Websuchdiensten erkannt wird.

Websuchdienste können kein einheitliches Verfahren zur Erschließung von Strukturinformationen anwenden, da die Dokumente im WWW keine einheitliche Struktur aufweisen. Trotzdem bestehen Möglichkeiten aus der Struktur der Dokumente Informationen zu entnehmen. Informationen können z. B. aus Layout- oder Navigationselementen von Dokumenten abgeleitet werden [Lewa05, 59], indem die Position der Inhalte oder die verwendeten Umschreibungen in den Hypertext Markup Language (HTML) Tags [HMSi02, 9 f.] analysiert werden.

Inhalte und Umschreibungen von Dokumenten liegen im WWW in verschiedenen Sprachen oder unter Verwendung verschiedener Schriftzeichen vor und können nur dann gefunden werden, wenn die Suchanfrage in der entsprechenden Sprache formuliert wurde oder Übersetzungen zur Überwindung des Sprachproblems eingesetzt werden [GrFr04, 149].

Die Qualität der Dokumenteninhalte im WWW ist heterogen, da das WWW sowohl qualifizierte als auch unverständlich geschriebene Texte oder Texte mit Rechtschreibfehlern enthält [BaRi99, 368]. Darüber hinaus können auch Dokumente mit falschen oder veralteten Informationen nicht ausgeschlossen werden [HMSi02, 5]. Erfolgt die Erschließung dieser Inhalte durch automatische Verfahren, müssen diese Verfahren auch auf Kriterien zur Qualitätsbewertung der Inhalte zurückgreifen können.³⁴

Die Aktualität der Dokumente im WWW ist für Websuchdienste ebenfalls schwer bestimmbar, da sich im WWW ein hoher Prozentsatz an un stetigen Daten befindet [BaRi99, 368]. Dieses können neue, verschobene, veränderte oder gelöschte Dokumente sein [Lewa05, 50]. Da das Datum des Dokuments nicht immer eine inhaltliche Änderung widerspiegeln muss, ist die Bestimmung der Aktualität der enthaltenen Informationen problematisch [Ferb03, 288]. Eine ständige Kontrolle der Verfügbarkeit und Aktualität der Dokumente ist auf Grund der Größe des WWW (bisher) nicht umsetzbar.

³⁴ Ein Kriterium ergibt sich z. B. aus der Auswertung von Verlinkungen im WWW auf diese Inhalte [Lewa05, 117].

Neben Dokumenten mit heterogenen Eigenschaften kann das WWW auch Dokumente mit gleichen oder sehr ähnlichen Inhalten enthalten.³⁵ Diese Dubletten stellen keinen Mehrwert für den Nutzer dar und sollten deshalb von Websuchdiensten ausgeschlossen werden.

Ebenso ist Spam ein unerwünschter Inhalt. Spam wird in diesem Zusammenhang als die Manipulation der Suchergebnisse von Websuchdiensten umschrieben [HMSi02, 2]. Henzinger, Motwani und Silverstein unterscheiden nach drei Arten von Spam: Text Spam, Link Spam und Cloaking.³⁶ Text Spam bezeichnet das Modifizieren eines Textes, so dass Websuchdienste diesem eine höhere Relevanz zuschreiben, obwohl für den Nutzer kein Mehrwert erkennbar ist. Link Spam nutzt die Tatsache aus, dass Websuchdienste die Popularität von Dokumenten auf Grund der Anzahl und Gewichtung von Verlinkungen im WWW auswerten und dies in der Anzeige der Ergebnisreihenfolge berücksichtigen. Zur Manipulation der Ergebnisreihenfolge wird versucht, möglichst viele und hochwertige³⁷ Verlinkungen auf die eigenen Dokumente im WWW zu erstellen. Beim Cloaking werden den Websuchdiensten andere Inhalte präsentiert als beim Abruf des Dokuments durch einen Nutzer. Hierbei wird ausgenutzt, dass automatische Verfahren bestimmte Inhalte anders verarbeiten als menschliche Nutzer.

Gründe für Spam sind meist die finanziellen Interessen von Anbietern. Im WWW hat sich deshalb ein Geschäftsfeld, das Suchmaschinenmarketing (SEM), entwickelt, welches sich mit den Maßnahmen zur Erhöhung der Sichtbarkeit³⁸ des eigenen Angebots in Websuchdiensten beschäftigt [GBRi09, 26]. Während das SEM auch versucht, durch gezielte Anzeigen in Websuchdiensten die Sichtbarkeit zu erhöhen, befasst sich der Teilbereich Suchmaschinenoptimierung (SEO) ausschließlich mit der Verbesserung der Sichtbarkeit in den regulären Trefferlisten der Websuchdienste [Lewa08, 7]. Spam, SEM und SEO können für Nutzer von Websuchdiensten zu unerwünschten und beeinflussten Ergebnissen bei der Websuche führen.

³⁵ Vgl. zum folgenden Abschnitt [Lewa05, 72 f.].

³⁶ Vgl. zum folgenden Abschnitt [HMSi02, 3 f.].

³⁷ Hochwertige Verlinkungen stammen von Webseiten, die über eine gewisse Reputation verfügen, z. B. in Bezug auf ein bestimmtes Thema.

³⁸ Als Sichtbarkeit wird hier die Wahrnehmung von Webseiten durch Nutzer und Websuchdienste verstanden.

2.2.3 Nutzer von Websuchdiensten

In Bezug auf die Nutzer von Websuchdiensten vergleicht Lewandowski verschiedene Nutzerstudien miteinander und fasst die Ergebnisse wie folgt zusammen.³⁹ Er kommt zu dem Schluss, dass die Nutzer sehr heterogene Interessen haben und deshalb die Suchanfragen an Websuchdienste sehr heterogen sind. Zudem wird festgestellt, dass die Nutzer Probleme bei der Übersetzung ihres Informationsbedürfnisses in eine Suchanfrage haben und nur geringe Kenntnisse über die Suchmöglichkeiten sowie den Suchprozess der Websuchdienste besitzen. Weitere Untersuchungen zum Nutzerverhalten ließen sich aus den Daten der Log Files⁴⁰ der Suchdiensteanbieter ableiten [CrHa09, 87]. Diese Log Files werden allerdings aus kommerziellen oder datenschutzrechtlichen Gründen nicht veröffentlicht.

2.2.4 Schlussfolgerungen

Das WWW besitzt für Websuchdienste relevante Eigenschaften, die sich in Bezug auf die Struktur und die Daten des WWW unterscheiden lassen. Diese Eigenschaften können zum einen Probleme erzeugen, z. B. bei der Erschließung der Inhalte oder bei der Qualitäts- und Relevanzbewertung der Ergebnisse der Websuchdienste; zum anderen aber auch von Nutzen sein, z. B. durch Zusatzinformationen aus Strukturelementen, die wiederum auch aus kommerziellen Interessen beeinflusst werden können. Bezüglich der Nutzer von Websuchdiensten sollte von geringeren Vorkenntnissen und geringeren Fertigkeiten in dem Umgang mit Websuchdiensten ausgegangen werden, als bei Nutzern von Verfahren des klassischen IR. Die Erkenntnisse dieses Abschnitts sollten bei den Ausführungen zu den Klassifizierungsansätzen und Kategorien von alternativen Websuchdiensten in Kapitel drei berücksichtigt werden.

2.3 Markt für Websuchdienste

Um einen Überblick über die Bedeutung alternativer Websuchdienste in Bezug auf die Gesamtheit an Websuchdiensten zu erhalten, wird in diesem Abschnitt auf die Stellung der alternativen Websuchdienste auf dem Markt für Websuchdienste eingegangen.

³⁹ Vgl. zum folgenden Abschnitt [Lewa05, 33 ff.].

⁴⁰ In Log Files werden von den Websuchdiensten die Suchanfragen und Auswahlentscheidungen der Nutzer gespeichert und zur Verbesserung des eigenen Websuchdienstes verwendet [Höch09, 180 f.].

2.3.1 Verteilung der Anteile

Für den Markt der Websuchdienste ist der überragende Marktanteil des Anbieters Google charakteristisch. In Deutschland werden fast 90% der Suchanfragen über Google ausgeführt (siehe Abb. 2).⁴¹ In den USA fällt der Anteil für Google zwar etwas niedriger aus, jedoch ist durch einen Anteil von 65,1% ebenfalls eine dominierende Stellung festzustellen (siehe Abb. 3). Des Weiteren sind auf der Abb. 3 des U.S. Marktes die zwei stärksten Konkurrenten von Google erkennbar. Microsofts Suchmaschine Bing und der Suchmaschinenanbieter Yahoo liegen mit einem Anteil von 13,9% bzw. 13,1% auf den Plätzen zwei und drei. Damit entfällt sowohl auf dem U.S. Markt als auch auf dem deutschen Markt fast der gesamte Marktanteil auf die Anbieter Google, Microsoft und Yahoo.

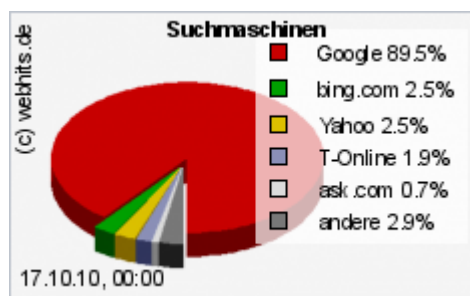


Abb. 2: Nutzung von Websuchdiensten auf dem deutschen Markt⁴²

Top 10 U.S. Search Providers, Home & Work

August 2010

| RANK | PROVIDER | SEARCHES (000) | SHARE OF SEARCHES |
|------|------------------------------|----------------|-------------------|
| - | ALL SEARCH | 9,199,567 | 100.0% |
| 1 | GOOGLE SEARCH | 5,988,996 | 65.1% |
| 2 | MSN/WINDOWS LIVE/BING SEARCH | 1,274,184 | 13.9% |
| 3 | YAHOO! SEARCH | 1,208,774 | 13.1% |
| 4 | ASK.COM SEARCH | 196,875 | 2.1% |
| 5 | AOL SEARCH | 179,895 | 2.0% |

Source: Nielsen MegaView Search

Abb. 3: Nutzung von Websuchdiensten auf dem U.S. Markt⁴³

⁴¹ In der Abb. 2 wird der Begriff Suchmaschinen repräsentativ für Websuchdienste verwendet.

⁴² <http://www.webhits.de/>, Abruf: 2010-10-17

⁴³ http://en-us.nielsen.com/content/nielsen/en_us/insights/rankings/internet.html, Abruf: 2010-10-17

2.3.2 Entwicklung

Wie aus Abb. 4 hervorgeht, hat Google, obwohl das Unternehmen erst 1998 in den Markt eingestiegen ist, seit 2002 einen Marktanteil von über 50% in Deutschland.

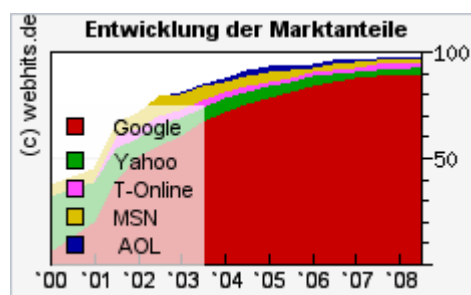


Abb. 4: Entwicklung der Marktanteile auf dem deutschen Markt⁴⁴

Der Aufstieg von Google wird der Einführung eines neuen Ansatzes zur Bewertung der Qualität der Treffer zugeschrieben (PageRank) [MSHP09, 5]. Demgegenüber hat der Anbieter Yahoo, der durch sein Verzeichnis bekannt geworden ist, seit 2001 große Anteile verloren. Dies kann auf die zunehmende Beliebtheit der Suchmaschinen zurückgeführt werden. Yahoo nahm daraufhin ebenso eine Suchmaschine in ihr Webangebot auf, allerdings mit zugekauften Ergebnissen [GBRi09, 21]. Erst später wurde mit dem Aufbau eines eigenen Index begonnen. Im Jahr 2010 sind Yahoo und Bing eine Kooperation eingegangen (Search Alliance).⁴⁵ Seitdem werden die Suchergebnisse von Yahoo aus dem Index von Bing generiert.

Trotz der hohen Marktanteile der wenigen großen Anbieter existiert auf dem Websuchdienstmarkt eine Vielzahl von kleinen Websuchdiensteanbietern.⁴⁶ Die großen Anbieter sind jedoch eine der wenigen Websuchdiensteanbieter mit eigenem Index.⁴⁷ Die weiteren Anbieter von Websuchdiensten beziehen ihre Ergebnisse häufig aus dem Index eines großen Anbieters. Der geringe Bestand an Anbietern mit eigenem Index ist zum einen mit Zusammenschlüssen zwischen den Anbietern bzw. Übernahmen durch die großen

⁴⁴ <http://www.webhits.de/>, Abruf: 2010-10-17

⁴⁵ Vgl. <http://www.searchalliance.com/home>, Abruf: 2010-09-22

⁴⁶ Vgl. z. B. <http://www.searchenginecolossus.com/>, <http://www.theseengineinelist.com/>, <http://www.suchlexikon.de/> oder <http://www.klug-suchen.de>, Abruf: 2010-08-05

⁴⁷ Vgl. dazu die Abbildung des Beziehungsgeflechts der Suchmaschinen unter <http://www.lunapark.de/home/internet-fakten/suchmaschinen-beziehungsgeflecht.html>, Abruf: 2010-10-17

Anbieter und zum anderen durch die hohen Kosten für den Betrieb der Suche mit eigenem Index sowie Investitionen in die Suchtechnologie zu erklären.

2.3.3 Chancen für alternative Websuchdienste

Um Aussagen über die Chancen für alternative Websuchdienste auf dem Websuchdienstmarkt treffen zu können, muss zuerst die derzeitige Situation analysiert werden.

Maaß et al. schätzen die Dominanz der drei großen Suchmaschinenanbieter als ungefährdet ein.⁴⁸ Nach ihrer industrieökonomischen Untersuchung des Marktes kamen diese zu dem Schluss, dass durch

- die geringe Bedrohung durch potenzielle neue Konkurrenten,
- der geringen Verhandlungsstärke der Lieferanten und Abnehmer,
- der hohen Rivalität unter den bestehenden Wettbewerbern in der Branche
- und der geringen Bedrohung durch Ersatzprodukte

kurz- und mittelfristig keine Änderungen an der Aufteilung des Marktes zu erwarten sind.

Eine geringe Bedrohung durch potenzielle neue Konkurrenten wird mit hohen Markteintrittsbarrieren begründet. Um in den Markt eintreten zu können, müssen Anbieter das technische Wissen und die finanziellen Mittel für den Betrieb eines Suchangebots besitzen. Kosten entstehen dabei zum einen durch den Umgang mit Daten, die der Größe und der Dynamik des WWW unterliegen, und zum anderen durch die Überbrückung des Zeitraums bis ein neuer Anbieter attraktiv genug für den Werbemarkt ist und daraus seine Einnahmen beziehen kann.

Die Lieferanten und Abnehmer auf dem Websuchdienstmarkt sind die Anbieter von Informationen bzw. die Nutzer von Websuchdiensten. Beide besitzen gegenüber den Websuchdienstanbietern eine geringe Verhandlungsstärke. Während die Anbieter von Informationen auf eine hohe Sichtbarkeit ihres Angebots in den Websuchdiensten angewiesen sind, damit ihr Angebot von möglichst vielen Personen gefunden wird, stehen den Nutzern der Websuchdienste noch zu wenige qualitativ gleichwertige Alternativen zur Verfügung oder sind sogar an bestimmte Websuchdienste gebunden⁴⁹.

⁴⁸ Die Ausführungen der folgenden Absätze sind [MSHP09, 9 ff.] entnommen worden.

⁴⁹ Eine Bindung an einen Websuchdienst kann z. B. durch die Erstellung eines Benutzerkontos zur Nutzung von Zusatzfunktionen entstehen. Zusatzfunktion ist z. B. die Anpassung der Suchfunktion an das

Die Rivalität unter den bestehenden Wettbewerbern ist hoch. Hohe Investitionen werden von den bestehenden Anbietern in den Betrieb der Suche und in die Forschung getätigt, um den Anschluss zur Konkurrenz nicht zu verlieren.

Durch die genannten Ausprägungen der Wettbewerbskräfte lassen sich folgende Empfehlungen für alternative Websuchdienste festhalten. Anbieter eines Websuchdienstes sollten sich nicht als direkter Konkurrent zu den allgemeinen Suchmaschinen auf dem Markt der Websuchdienste platzieren, um nicht an der hohen Wettbewerbsintensität zu scheitern. Dieses Risiko kann durch das Anbieten eines alternativen Websuchdienstes verringert werden. Weiter sollte die Qualität der Suchergebnisse von alternativen Websuchdiensten mindestens der Qualität der Suchergebnisse der allgemeinen Suchmaschinen entsprechen. Zudem sollte ein Bekanntheitsgrad angestrebt werden, mit dem sich Einnahmen aus dem Werbemarkt generieren lassen.

Diese Hinweise haben die drei großen Suchmaschinenanbieter Google, Microsoft und Yahoo in ihren Angeboten bereits berücksichtigt. Alle drei bieten eine Vielzahl von Spezialsuchmaschinen an, können bei der Suche zusätzlich auf qualitative Ergebnisse aus der allgemeinen Suche zurückgreifen und besitzen einen Bekanntheitsgrad, der sowohl bei den Nutzern als auch bei den Werbetreibenden für hohe Aufmerksamkeit sorgt. Daraus wird deutlich, dass die großen Anbieter ihre starke Marktposition gegenüber Anbietern mit neuen Suchansätzen absichern können und ebenfalls in neue Suchansätze investieren.

2.3.4 Schlussfolgerungen

In Bezug auf die Marktanteile haben alternative Websuchdienste derzeit kaum eine Bedeutung im Vergleich zu den allgemeinen Suchmaschinen. Jedoch können sich alternative Websuchdienste abseits der allgemeinen Suchmaschinen auf dem Markt positionieren. Somit können alternative Websuchdienste mit neuen oder verbesserten Suchansätzen für eine Weiterentwicklung der Websuchdienste sorgen.⁵⁰ Die Tätigkeitsfelder der Anbieter Google, Microsoft und Yahoo sind allerdings so weit ausgebaut, dass die Ideen neuer, alternativer Suchansätze auch von den drei großen Anbietern aufgenommen werden und die Lücken in ihrem Angebot von Suchfunktionen durch die Integration eigener Alternativen geschlossen werden. Deshalb ist es für kleine Anbieter von

individuelle Suchverhalten. Vgl. Abschnitt 3.1.9

⁵⁰ Die Einführung des PageRank durch Google ist ein Beispiel für die Weiterentwicklung der Websuchdienste durch einen verbesserten Suchansatz.

alternativen Websuchdiensten schwer, sich auf dem Markt für Websuchdienste zu behaupten und eine Veränderung der Marktsituation herbeizuführen.

2.4 Kennzahlen für einen Vergleich von Websuchdiensten

In diesem Abschnitt werden die Grundlagen zu den verwendeten Kennzahlen für den Vergleich alternativer Websuchdienste in Kapitel vier erläutert.

Die Wahl der Kennzahlen für einen Vergleich von Websuchdiensten ist abhängig von den Eigenschaften der Websuchdienste, die verglichen werden sollen. Für einen Vergleich der Inhalte bzw. der Qualität von Websuchdiensten können klassische Kennzahlen zur Retrievaleffektivität (z. B. Precision⁵¹ oder Recall⁵²), web-spezifische Vergleichsgrößen (z.B. Indexgröße und -aktualität) oder Untersuchungen in Bezug auf die Nutzer der Websuchdienste (z. B. Bedienbarkeit und Bedürfnisbefriedigung) dienen [LeHö07; LaMe06, 8]. Gegenstand dieser Arbeit ist allerdings ein Vergleich von alternativen Websuchdiensten anhand der Popularität. Dazu kann auf Kennzahlen zurückgegriffen werden, die auf Annahmen aus dem Forschungsbereich der Webometrie basieren.

Die Webometrie ist ein Teilbereich der Informationswissenschaften⁵³ und wurde aus der Bibliometrie [Prit69] abgeleitet [Alln97]. In der Webometrie werden die Struktur, Nutzung und Inhalte des WWW untersucht. Björneborn definierte den Begriff Webometrie im Jahr 2004 als „the study of the quantitative aspects of the construction and use of information resources, structures and technologies on the Web drawing on bibliometric and informetric approaches“ [Björ04, 12]. Aus einer Definition von Thelwall aus dem Jahr 2009 lässt sich entnehmen, dass die Webometrie auch für Untersuchungen außerhalb der Informationswissenschaften genutzt werden kann: „the study of web-based content with primarily quantitative methods for social science research goals using techniques that are not specific to one field of study“ [Thel09, 6].

Die Ableitung der Webometrie aus der Bibliometrie ist auf die Erkennung der strukturellen Ähnlichkeit von Links und Zitationen zurückzuführen [Lars96; Rous97] Während in der Bibliometrie in Zitationsanalysen die Zitate als Referenzen zwischen wissenschaftlichen

⁵¹ Die Kennzahl Precision beschreibt das Verhältnis von der Anzahl relevanter Dokumente, die gefunden wurden, zu der Anzahl aller gefundenen Dokumenten [Lewa05, 140].

⁵² Die Kennzahl Recall beschreibt das Verhältnis von der Anzahl relevanter Dokumente, die gefunden wurden, zu der Anzahl aller relevanten Dokumente [Lewa05, 139].

⁵³ Zur Abgrenzung der Webometrie von anderen Forschungsbereichen vgl. [BjIn04, 1217].

Texten ausgewertet werden [Garf79], werden in der Webometrie in Linkanalysen die Verlinkungen zwischen Webseiten als Empfehlungen ausgewertet [Alln97]. Für Linkanalysen ist somit die Annahme, „dass sich die Außenwirkung veröffentlichter Informationen durch Links – ebenso wie durch Zitationen – messen lässt“ [ThRu06, 401], grundlegend. Durch diesen strukturzentrierten Ursprung der Webometrie wurden bisher zwar überwiegend strukturelle Eigenschaften des WWW untersucht, die Analyse von Nutzung und Inhalten des WWW gehört jedoch ebenfalls dem Forschungsbereich der Webometrie an [Vaug08, 1].

Die Kennzahlen, die in webometrischen Untersuchungen und in dieser Arbeit verwendet werden können, lassen sich in drei Kategorien einteilen. Lo und Sedhain stellen diese Kategorien als Kriterien dar, auf denen Methoden basieren, die zur Erstellung eines Rankings von Websites verwendet werden.⁵⁴ Die Methode, mit der in Kapitel vier der Vergleich alternativer Websuchdienste durchgeführt wird, wertet ebenfalls Webseiten⁵⁵ aus, da die alternativen Websuchdienste über Webseiten im WWW bereitgestellt werden. Lo und Sedhain unterscheiden zwischen referenzbezogenen Kriterien, aktivitätsbezogenen Kriterien und meinungsbezogenen Kriterien.

Mit referenzbezogenen Kriterien können Webseiten anhand der Anzahl und Gewichtung der Verlinkungen auf diese Webseiten verglichen werden. Die Gewichtung einer Verlinkung ist hoch, wenn die Verlinkung von einer einflussreichen Webseite stammt. Der Einfluss einer Webseite ergibt sich wiederum aus der Anzahl der Verlinkungen auf diese Webseite oder aus der Reputation der Webseite in Bezug auf ein bestimmtes Thema. Bekannte Kennzahlen, die solche referenzbezogenen Kriterien auswerten, sind z. B. der WIF und der PageRank.

Mit aktivitätsbezogenen Kriterien können Webseiten anhand der Anzahl von Aktivitäten auf den einzelnen Webseiten verglichen werden. Eine bekannte Kennzahl für aktivitätsbezogene Kriterien ist z. B. der Traffic (Datenverkehr) von Webseiten. Daten für solche Kennzahlen können im WWW auf der Anbieterseite und auf der Nutzerseite gesammelt werden. Auf der Anbieterseite werden die Daten auf dem Web Server (z. B. die

⁵⁴ Die Ausführungen der folgenden Absätze sind [LoSe06, 233 f.] entnommen worden.

⁵⁵ Die Unterscheidung zwischen den Begriffen Website und Webseite ist für die Erläuterung der folgenden Kriterien nicht relevant.

Visits⁵⁶ oder Page Views⁵⁷) erfasst. Auf der Nutzerseite werden die Daten durch den Browser des Nutzers (z. B. durch Cookies⁵⁸ oder Registrierungen) aufgezeichnet.

Mit meinungsbezogenen Kriterien können Webseiten anhand subjektiver Beurteilungen zu diesen Webseiten verglichen werden. In diesem Fall basieren die Bewertungen der Webseiten überwiegend auf Meinungen von einzelnen Personen oder kleinen Gruppen, die einem bestimmten Interessenbereich oder einer Redaktion zu einem bestimmten Thema angehören.

In dieser Arbeit wird der Vergleich alternativer Websuchdienste mit der Hilfe von linktopologischen Kennzahlen und Nutzungskennzahlen durchgeführt. Die linktopologischen Kennzahlen werten die Webseiten der Websuchdienste nach referenzbezogenen Kriterien aus und die Nutzungskennzahlen nach aktivitätsbezogenen Kriterien. Meinungsbezogene Kriterien liegen in einem zu geringen Umfang vor oder lassen sich auf Grund unterschiedlicher Bewertungskriterien und Erstellungszeitpunkte nicht aggregieren.

2.4.1 Linktopologische Kennzahlen

Linktopologische Kennzahlen werden aus der Anzahl und Gewichtung von Verlinkungen von Webseiten im WWW berechnet. Für die Bewertung der Popularität einer Webseite sind linktopologische Kennzahlen geeignet, da das Verlinken einer Webseite im Allgemeinen als eine Empfehlung für das verlinkte Dokument gewertet werden kann.

Die Daten zur Berechnung der Kennzahlen können entweder bei Suchmaschinen und Linkdatenbanken abgerufen werden oder durch einen eigenen Crawler⁵⁹ erfasst werden [ThRu06, 401]. Auf Grund der Struktur des WWW (vgl. Abschnitt 2.2.1) kann sowohl durch einen eigenen Crawler als auch durch eine Suchmaschinen keine vollständige Erfassung des WWW erfolgen. Es kann allerdings davon ausgegangen werden, dass im Rahmen dieser Arbeit über eine Suchmaschine eine größere Abdeckung erreicht werden kann als mit einem eigenen Crawler, da die großen Suchmaschinenbetreiber über

⁵⁶ Als Visit wird der Besuch einer Website durch einen Nutzer bezeichnet. Ein Visit kann aus mehreren Webseitenaufrufen durch denselben Nutzer bestehen, wenn diese innerhalb eines bestimmten Zeitraumes liegen [IVWR09, 6].

⁵⁷ Als Page Views (auch Page Impressions) wird der Abruf von Webseiten innerhalb einer Website durch einen Nutzer bezeichnet [IVWR09, 4].

⁵⁸ Cookies werden von Webbrowsern auf dem Rechner eines Nutzers gespeichert und dienen Webservern zur Identifikation von Nutzern, um z. B. deren persönliche Einstellungen für Websites abzurufen.

⁵⁹ Als Crawler wird ein automatischer Prozess von Websuchdiensten bezeichnet, der nach algorithmischen Verfahren Webseiten im WWW und deren Inhalte erschließt [Lewa05, 27].

leistungsstarke Hardware und Erfahrungen bezüglich der Erfassung des WWW verfügen. Zudem nutzen diese neben den eigenen automatischen Verfahren auch manuelle Möglichkeiten zur Erfassung des WWW, wie z. B. die manuelle Eintragung von Webseiten durch Nutzer.⁶⁰

Neben der unvollständigen Abdeckung bei der Erfassung von Webseiten und deren Verlinkungen ergibt sich eine weitere Unsicherheit. Das Verlinken einer Webseite muss nicht immer einer Empfehlung der Webseite entsprechen. So kann eine Verlinkung z. B. auch zur Anführung eines negativen Beispiels erstellt werden [Lewa04b, 11]. Einige Nutzer geben Verlinkungen auch nur zur Erhöhung der Sichtbarkeit der eigenen Webseiten an [Dopi09, 111], da die Anzahl und der Wert der Verlinkungen auf Webseiten zwei wesentliche Faktoren für die Ergebnisrankings der allgemeinen Suchmaschinen sind [Lewa04b, 1].

Diese und weitere Störfaktoren sollten bei der Durchführung und Auswertung von Linkanalysen beachtet werden.⁶¹ Trotz der genannten Unsicherheiten kann die Anzahl und die Gewichtung von Verlinkungen für die Bewertung der Popularität von Websuchdiensten ausgewertet werden, da bisherige Linkanalysen in verschiedenen Wissenschaftsbereichen angewendet wurden und zu nützlichen Ergebnissen geführt haben. Für den Vergleich alternativer Websuchdienste wurden die linktopologischen Kennzahlen WIF und PageRank ausgewählt, da diese über einen hohen Bekanntheitsgrad im Zusammenhang mit webometrischen Untersuchungen verfügen und mit einfachen Mitteln abgerufen werden können.

2.4.1.1 Web Impact Factor

Der WIF entstand aus der Übertragung des Impact Factors (IF) der Bibliometrie auf das WWW.⁶² In der Webometrie wird mit dem WIF die Attraktivität bzw. die Außenwirkung [ThRu06, 401] einer Website zu einem bestimmten Zeitpunkt gemessen. Der WIF kann für unterschiedliche Untersuchungen variiert werden, so dass verschiedene, für die Untersuchung relevante Faktoren (z. B. Anzahl der Webseiten einer Website, Anzahl der Mitarbeiter einer Universität) bei der Berechnung berücksichtigt werden [The109, 2 ff.]. Der einfache WIF nach Ingwersen errechnet sich aus der Summe der Verlinkungen auf eine

⁶⁰ Vgl. <http://www.google.de/addurl/>, Abruf: 2010-08-03

⁶¹ Vgl. zum folgenden Abschnitt [ThRu06, 401].

⁶² Vgl. zum folgenden Abschnitt [Ingw98, 236 ff.].

Website, dividiert durch die Anzahl der Webseiten, aus denen die Website besteht [Ingw98, 237]. Im Allgemeinen gilt also, dass bei einem Vergleich von Websites, gemessen an dem WIF, eine Website mit dem größeren WIF eine höhere Popularität besitzt als eine Website mit einem kleineren WIF.

2.4.1.2 PageRank

Der PageRank wurde von Brin und Page zur Verbesserung der Ergebnisrankings von Suchmaschinen entwickelt [BrPa98, o. S.]. Seitdem ist die Idee des PageRanks Grundlage für eine Vielzahl von Suchmaschinen [LaMe06, 4]. Der Grundgedanke des PageRanks besteht darin, zusätzlich zu der Anzahl, auch eine Gewichtung der Verlinkungen von Webseiten in die Bewertung von Suchergebnissen mit einzubeziehen [PBMW98, 3]. Die Gewichtung einer Verlinkung bezieht sich darauf, dass Verlinkungen auf eine Webseite von Webseiten mit unterschiedlicher Reputation stammen können [Lewa04b, 1]. Von einer hohen Reputation einer Webseite kann ausgegangen werden, wenn diese Webseite wiederum viele Verlinkungen von anderen Webseiten erhält oder ein hohes Ansehen in Bezug auf ein bestimmtes Thema besitzt [Dopi09, 106]. Bei der Berechnung des PageRanks gibt eine Webseite den eigenen PageRank weiter, indem dieser gleichmäßig auf die verlinkten Webseiten verteilt wird.⁶³ Der PageRank einer Webseite ist also rekursiv definiert, da dieser aus der Bedeutsamkeit anderer Webseiten berechnet wird und wiederum Auswirkungen auf den PageRank anderer verlinkter Webseiten hat.

Im Allgemeinen gilt, dass bei einem Vergleich von Webseiten, gemessen an dem PageRank, eine Webseite mit einem höheren PageRank eine höhere Popularität besitzt als eine Webseite mit einem niedrigeren PageRank.

2.4.2 Nutzungskennzahlen

Für die Berechnung von Nutzungskennzahlen können neue Nutzungsdaten zu Webseiten der Websuchdienste erhoben oder vorhandene Nutzungsdaten verwendet werden.

Neue Nutzungsdaten können durch Befragungen, Verhaltensbeobachtungen, Experimente oder Simulationsstudien erhoben werden.⁶⁴ Diese Methoden sind mit einem hohem

⁶³ Vgl. zum folgenden Abschnitt [PBMW98 3 f.].

⁶⁴ Die Ausführungen der folgenden Absätze sind [Höch09] entnommen worden. Die Ausführungen von Höchstötter beziehen sich auf die Nutzung von Websuchdiensten (Suchmaschinen), lassen sich hier aber auch auf die Nutzung von Webseiten, auf denen sich die Websuchdienste befinden, übertragen.

Aufwand verbunden, erheben häufig nur eine geringe Menge an Daten von einer kleinen Anzahl von Nutzern und werden überwiegend unter Überwachung durchgeführt. Da den beobachteten Personen diese Überwachung bewusst ist, kann es vorkommen, dass die Ergebnisse durch die Überwachung beeinflusst werden.

Bestehende Nutzungsdaten befinden sich in den Log Files von Websuchdiensten oder lassen sich Webdiensten entnehmen, die die Nutzungsdaten über ein eigenes Online Panel ermitteln.

Die Log Files werden von den Websuchdienstbetreibern in der Regel nicht veröffentlicht und stehen deshalb nicht in dieser Arbeit zur Verfügung.

Mit einem Online Panel werden Nutzungsdaten durch eine spezielle Software automatisch, im Hintergrund und kontinuierlich aufgezeichnet. Obwohl das Sammeln von Daten durch Online Panels auch eine überwachte Methode darstellt, ist diese Überwachung dem Nutzer nicht so allgegenwärtig bewusst, wie bei den zuvor genannten Methoden durch die Anwesenheit einer überwachenden Person. Ein weiterer Vorteil ist die Vielfalt von Daten und die große Anzahl von Nutzern, die ausgewertet werden können, da für die Nutzer, bis auf die einmalige Einrichtung des Online Panels, kein zusätzlicher Aufwand entsteht. Zudem lassen sich auf Grund der kontinuierlichen Datenaufzeichnung durch Online Panels Veränderungen in den Nutzungsdaten beobachten. Die Daten aus Online Panels sind allerdings auch kritisch zu betrachten, da sich die Teilnehmer dafür überwiegend selbst auswählen und anmelden können. Außerdem steht es den Nutzern häufig frei, demografische Angaben zu tätigen. Bei der Verwendung dieser Daten ist deshalb zu beachten, dass die erhobenen Nutzungsdaten Nutzer repräsentieren, die sich selbst ausgewählt haben und personenbezogene Merkmale fehlen können.

Im WWW existieren verschiedene Webdienste, die Nutzungsdaten über ihr eigenes Online Panel in Form von einer Browsererweiterung, einer Toolbar, aufzeichnen. Beispiele für solche Webdienste sind Alexa Internet⁶⁵, Ranking.com⁶⁶ und WebSearch Webranking⁶⁷. Diese veröffentlichen die aufgezeichneten Daten direkt und/oder als berechnete Kennzahlen und erstellen daraus ein Ranking von Websites.⁶⁸

⁶⁵ <http://www.alexa.com/>, Abruf: 2010-09-10

⁶⁶ <http://www.ranking.com/>, Abruf: 2010-09-10

⁶⁷ <http://ranking.websearch.com/>, Abruf: 2010-09-10

⁶⁸ Vgl. <http://www.alexa.com/help/traffic-learn-more>, <http://www.ranking.com/about.htm>, <http://ranking.websearch.com/LearnMore.aspx>, Abruf: 2010-08-10

Alexa Internet ist der bekannteste Webdienst mit eigener Aufzeichnung von internationalen Nutzungsdaten im WWW und wird in dieser Arbeit zum Abruf von einer Nutzungskennzahl, dem Alexa Traffic Rank genutzt. Der Alexa Traffic Rank gibt an, welchen Rang eine Website in dem Ranking von Alexa Internet belegt. Für das Ranking wertet Alexa Internet sowohl die Anzahl von Besuchern einer Website (diese Anzahl wird bei Alexa Internet als Reach bezeichnet) als auch die Anzahl von Seitenaufrufen innerhalb einer Website (Page Views) aus.⁶⁹ Dabei werden mehrfache Aufrufe derselben Seite durch denselben Nutzer am selben Tag als ein Aufruf gewertet. Für den Alexa Traffic Rank wird das arithmetische Mittel der Werte von Reach und Page Views berechnet, die innerhalb der letzten drei Monate gemessen wurden.⁷⁰ Alexa Internet weist allerdings darauf hin, dass durch die Daten keine Repräsentation aller Internetnutzer gewährleistet werden kann, da die Daten überwiegend von Nutzern stammen, die die Alexa Toolbar installiert haben. Um potentiellen Verzerrungen entgegenzuwirken, versucht Alexa Internet bei dem Ranking der Websites Methoden anzuwenden, die auch Website-Besucher berücksichtigen, die nicht die Alexa Toolbar nutzen. Diese Methoden werden allerdings nicht weiter von Alexa Internet erläutert. Zudem gibt Alexa Internet an, dass es durch die Größe des WWW und den automatischen Verfahren von Alexa Internet, besonders bei Websites mit weniger als 1000 Besuchern im Monat, zu Ungenauigkeiten bei dem Ranking der Websites kommen kann.

Trotz dieser Unsicherheiten kann der Alexa Traffic Rank für den Vergleich von alternativen Websuchdiensten verwendet werden, da diese Kennzahl bereits in einer weiteren webometrischen Untersuchung bestätigt wurde⁷¹ und das beste zur Verfügung stehende Mittel für den Abruf von Nutzungskennzahlen in dieser Arbeit darstellt. Im Allgemeinen gilt, dass bei einem Vergleich von Websites, gemessen an dem Alexa Traffic Rank, eine Website mit einem höheren Rang eine höhere Popularität besitzt als eine Website mit einem niedrigeren Rang.

⁶⁹ Vgl. zum folgenden Abschnitt http://www.alexa.com/help/traffic_learn_more, Abruf: 2010-09-14

⁷⁰ Vgl. <http://blog.alexa.com/search?updated-max=2010-04-15T16%3A00%3A00-07%3A00&max-results=7>, Abruf: 2010-09-14

⁷¹ Vaughan konnte in einer webometrischen Untersuchung von Universitäten feststellen, dass eine signifikante Korrelation zwischen einem auf Nutzungsdaten basierenden Ranking von Alexa Internet und dem Ranking, das von einer anderen Institution aufgestellt wurde, besteht [Vaug08].

2.4.3 Schlussfolgerungen

Für den Vergleich der Popularität von alternativen Websuchdiensten werden Kennzahlen abgerufen, die die Verlinkungen und Nutzungsdaten der alternativen Webseiten der Websuchdienste auswerten.

Die Anzahl der Verlinkungen von Webseiten wird durch den WIF beschrieben, die Anzahl und Gewichtung der Verlinkungen durch den PageRank.

Für die Auswertung von Nutzungsdaten erweist sich der Rückgriff auf bestehende Daten als sinnvoll, weil dadurch eine größere Menge an Daten, eine größere Anzahl und breitere Verteilung von Nutzern berücksichtigt werden kann als dieses durch eine eigene Erhebung im Rahmen dieser Arbeit möglich wäre. Zudem kann eine eigene überwachte Aufzeichnung von Daten zu beeinflussten Ergebnissen führen. Deshalb werden die Nutzungsdaten zu den alternativen Websuchdiensten in den Vergleich mit einbezogen, indem deren Alexa Traffic Rank bei dem Webdienst von Alexa Internet abgerufen wird.

Die in diesem Abschnitt genannten Kennzahlen und Ausführungen werden in Kapitel vier bei der Durchführung des Vergleichs angewendet und bei der Interpretation der Rankings berücksichtigt.

3 Überblick über Klassifizierungsansätze und Kategorien

3.1 Klassifizierungsansätze

Damit ein Vergleich alternativer Websuchdienste innerhalb von Kategorien durchgeführt werden kann, muss zuerst abgeleitet werden, anhand welcher Merkmale Websuchdienste unterschieden werden können. Dazu werden im Folgenden die verschiedenen Suchansätze von Websuchdiensten dargestellt.

3.1.1 Nutzerführung

Der Nutzer kann durch Websuchdienste auf unterschiedliche Weise zur gewünschten Information geführt werden. Zum einen kann der Nutzer von der Startseite eines Websuchdienstes aus durch verschiedene, thematisch geordnete Kategorien navigiert

werden und dort einsortierte Ergebnisse auffinden⁷². Zum anderen können Websuchdienste ein Suchfeld anbieten, in welches der Nutzer eine Suchanfrage eingibt und daraufhin entsprechende Suchergebnisse erhält. Eine Kombination dieser Suchansätze ist ebenfalls möglich.⁷³ Das Suchergebnis selbst kann entweder eine direkte Antwort (z. B. eine Definition oder eine Zahl) sein [Lewa08, 2] oder es werden Verlinkungen auf andere Dokumente wiedergegeben, welche die gewünschte Information enthalten können. Die Reihenfolge der angezeigten Ergebnisse kann dabei durch verschiedene vom Websuchdienst ausgewertete Faktoren bestimmt sein [GBRi09, 33 f.]. Websuchdienste sind also nicht nur dazu da, das Informationsbedürfnis mit einer Antwort auf die Suchanfrage zu befriedigen, sondern können ebenso zur Navigation und Erkundung des WWW genutzt werden oder zum Erreichen einer Webseite dienen, auf denen der Nutzer weitere Aktionen oder Transaktionen durchführen möchte [Brod02, o. S.].

3.1.2 Allgemeine und spezielle Suche

Suchansätze können sich bezüglich ihrer Ausrichtung auf spezielle Informationsbedürfnisse der Nutzer unterscheiden. Zum einen können Websuchdienste eine allgemeine Ausrichtung verfolgen. In diesem Fall versuchen Websuchdienste möglichst alle Inhalte des WWW zu erfassen und passen ihre Suchfunktionen und Darstellung der Suchergebnisse an den Informationsbedürfnissen der Mehrheit der Nutzer an, um eine große Nutzermasse zu befriedigen (vgl. Abschnitt 2.2.3). Deshalb wird von allgemeinen Suchansätzen in besonders starkem Maße die Popularität von Dokumenten in den Rankings ihrer Suchergebnisse berücksichtigt [Lewa09c, 55].

Zum anderen können Websuchdienste auf die Beantwortung von speziellen Suchanfragen ausgerichtet sein. Dabei kann die Spezialität des Informationsbedürfnisses des Nutzers auf bestimmte Themen, Dokumenttypen oder -formate oder auf einen bestimmten geographischen oder sprachlichen Raum bezogen sein [Gele03, o. S.]. Dementsprechend ist auch die Erschließung der Inhalte durch den Websuchdienst auf den eingeschränkten Suchraum abgestimmt (vgl. Abschnitt 3.2). Zudem können dem Nutzer, je nach Themenbereich oder Dokumenttyp, spezielle Funktionen für die Suchanfrage bereitgestellt werden und die Anzeige der Ergebnisse durch den Websuchdienst an das Thema oder den Dokumenttyp angepasst werden [Lewa09c, 56].

⁷² Diese Aktivität wird auch als Browsing bezeichnet [BaRi99, 384 ff.].

⁷³ Siehe z. B. <http://www.dmoz.de/>, Abruf: 2010-08-28

3.1.3 Manuelle und automatische Inhaltserschließung

In Bezug auf die Erschließung der Inhalte im WWW lassen sich automatische und manuelle Verfahren unterscheiden.⁷⁴ Während bei der automatischen Erschließung mit der Hilfe von algorithmischen Verfahren ein Index erstellt wird, werden bei der manuellen Erschließung die Inhalte durch Personen, die überwiegend einer Redaktion oder Interessengruppe zu einem bestimmten Thema (Community) angehören, ausgewählt, umschrieben und in ein System eingetragen. Ebenso existieren Websuchdienste, die keinen eigenen Index erstellen, sondern den Index anderer Websuchdienste nutzen und daraus ihre eigenen Treffermengen generieren.⁷⁵

3.1.4 Deep Web-Suche

Der Begriff Deep Web bezeichnet den Teil des WWW, der auf Grund technischer Hindernisse durch allgemeine Suchmaschinen nicht erschlossen werden kann [ShPr01, 57]. Einen großen Teil⁷⁶ des Deep Web stellen Inhalte dar, die sich in Datenbanken befinden [Lewa05, 53 f.]. Dabei handelt es sich häufig um Inhalte von hoher Qualität (z. B. wissenschaftliche Inhalte), deren Abruf kosten- oder registrierungspflichtig ist. Ein alternativer Suchansatz ist es, das Deep Web durchsuchbar zu machen, indem Kooperationen und gemeinschaftliche Ansätze mit Verlagen, Bibliotheken und weiteren Informationsanbietern entwickelt werden [LeMa06, 10 ff.]. Auf diese Weise können z. B. große Datenbanken kommerzieller Informationsanbieter in HTML-Seiten konvertiert werden, um diese für Websuchdienste indizierbar zu machen [LeMa06, 12]. Kombinieren Suchansätze die im Deep Web enthaltenen Daten mit Ergebnissen aus dem Oberflächenweb, wird dies mit dem Begriff Hybridsuche beschrieben [Lewa09c, 58].

3.1.5 Soziale Suche

Eine Vielzahl alternativer Websuchdienste ist eng mit den Entwicklungen des Web 2.0 [O'Rei05, o. S.] verbunden. Das Web 2.0 ist dadurch gekennzeichnet, dass Technologien bereitgestellt werden, die es einer Vielzahl von Nutzern ohne großes technisches Vorwissen ermöglichen, an der Erstellung der Inhalte im WWW mitzuwirken [Heck08, 21 f.]. Die

⁷⁴ Vgl. zum folgenden Abschnitt z. B. [Lewa05, 24 ff.; BaRi99, 367; Ferb03, 294 ff.].

⁷⁵ Aggregiert ein Suchdienst seine Suchergebnisse aus den Indizes mehrerer Suchdienste, wird dieser Ansatz mit dem Begriff Metasuche beschrieben [Lewa05, 25].

⁷⁶ Für eine Übersicht der Typologie der Deep Web Inhalte siehe [ShPr01, 61].

Partizipation der Nutzer findet z. B. in sozialen Netzwerken, Blogs⁷⁷ oder Communities statt. Tim O'Reilly empfiehlt in seiner Beschreibung des Web 2.0, sich die „kollektive Intelligenz“ der Anwender nutzbar zu machen [O'Rei05, o. S.].

An dieser Stelle setzen Websuchdienste an, die mit dem Begriff Soziale Suche (Social Search) beschrieben werden. Nach Skusa und Maaß beziehen soziale Suchansätze das Wissen menschlicher Nutzer in den Suchprozess mit ein [SkMa08, 2]. Eine Form der Mitwirkung durch Nutzer stellt das Auszeichnen (Tagging) von Dokumenten bzw. von Verlinkungen auf Dokumente mittels Schlagwörtern (Tags) dar. Die Gesamtheit der Nutzer, die diese Verschlagwortung durchführen, wird als Tagging Community bezeichnet. Tagging Communities können ebenso auf bestimmte Bereiche des WWW begrenzt sein, wie z. B. auf Videos (YouTube⁷⁸) oder Bilder (Flickr⁷⁹). Die Tags selbst, deren Häufigkeit sowie deren Beziehungen untereinander können von verschiedenen Kategorien von alternativen Websuchdiensten verwertet werden (vgl. Abschnitt 3.2). Bezüglich der Auswirkungen von Tags auf Websuchdienste ist auf Lewandowski und Höchstötter zu verweisen: „Inwieweit dieser Ansatz zu Verbesserungen im Retrieval führt und/oder wie er sich am besten ergänzend zu den bekannten Verfahren einsetzen lässt, ist noch weitgehend unklar“ [LeHö08, 2]. Auf Grund der Freiheiten bei der Auszeichnung und der fehlenden Kontrolle entstehen durch Tagging Communities auch Probleme. Eine freie Formulierung der Tags hat zur Folge, dass die Suchanfrage mit den verwendeten Tags übereinstimmen muss, damit entsprechende Inhalte gefunden werden. Da die Auszeichnung von Inhalten durch Nutzer fehleranfällig ist (Rechtschreibfehler, falsche Synonymverwendung, Kombinationen von Tags), kommt es vor, dass Inhalte nicht oder im falschen Kontext wiedergefunden werden [SkMa08, 7]. Ebenso sind Tagging Communities anfällig für Manipulationen, da die verwendeten Tags nicht den Inhalten des Dokuments entsprechen müssen.

Neben den Tags können Websuchdienste auch Bewertungen und Kommentare von Nutzern im WWW auswerten.⁸⁰ Diese können in der Ergebnisreihenfolge und -anzeige der Websuchdienste direkt oder aggregiert berücksichtigt werden.

⁷⁷ „Blogs sind Webseiten, die zeitlich markierte Beiträge von einem oder mehreren Personen enthalten und in chronologisch entgegengesetzter Richtung sortiert sind“ [The109, 47].

⁷⁸ <http://www.youtube.com>, Abruf: 2010-08-17

⁷⁹ <http://www.flickr.com>, Abruf: 2010-08-17

⁸⁰ Vgl. zum folgenden Abschnitt [Lewa09b, 2 ff.].

3.1.6 Echtzeitsuche

Die Echtzeitsuche (Real Time Search) bezeichnet die Suche nach zeitkritischen Informationen.⁸¹ Zeitkritische Informationen müssen möglichst zeitnah erfasst werden, da deren Informationswert innerhalb kurzer Zeit abnimmt. Die Verfahren der allgemeinen Suchmaschinen, die zur Erreichung einer möglichst großen Abdeckung im WWW verwendet werden, sind (bisher) nicht geeignet, um eine zeitnahe Erfassung zu ermöglichen. Besonders durch die zunehmende Anzahl von Informationen, die durch eine große Anzahl von Nutzern mittels sozialer Netzwerke, Blogs und Twitter⁸² erstellt werden, hat die Echtzeitsuche aktuell stark an Bedeutung gewonnen. Geeignete Verfahren zur zeitnahen Erschließung solcher Informationen werden im Zusammenhang der alternativen Websuchdienstkategorien in Abschnitt 3.2.5 und 3.2.6 erläutert. Bisher ist die Echtzeitsuche jedoch auf ausgewählte Bereiche des WWW beschränkt.⁸³

3.1.7 Semantische Suche

Semantische Suchansätze basieren auf dem Verständnis eines Semantischen Webs, wie es von Tim Berners-Lee vorgestellt wurde.⁸⁴ Demnach soll das Semantische Web das existierende WWW erweitern, indem den Informationen im WWW durch zusätzliche Beschreibungen eine Bedeutung gegeben wird, die auch von Computern interpretiert und automatisch verarbeitet werden kann. Diese Beschreibungen sollten mit einer standardisierten Beschreibungssprache durchgeführt werden und, neben der Bedeutung von Begriffen, deren Beziehungen untereinander deutlich machen.⁸⁵ Durch die semantischen Auszeichnungen könnten Websuchdienste mit automatischen Verfahren zwischen gleichlautenden Begriffen mit unterschiedlicher Bedeutung unterscheiden und themenverwandte Informationen finden. Ebenso könnten die Beziehungen der Begriffe, die innerhalb einer komplexen Suchanfrage stehen, vom Websuchdienst berücksichtigt werden und exakte Antworten wiedergegeben werden.

⁸¹ Vgl. zum folgenden Abschnitt [Lewa09a, 1 f.].

⁸² Über den Webdienst Twitter können Nutzer Kurznachrichten im WWW veröffentlichen.
<http://twitter.com/>, Abruf: 2010-09-15

⁸³ Der Suchraum für die Echtzeitsuche beschränkt sich bei den meisten Websuchdiensten auf Blogs, Twitter und soziale Netzwerke.

⁸⁴ Vgl. zum folgenden Abschnitt [BHLa01, o. S.].

⁸⁵ Vgl. zum folgenden Abschnitt [SkMa08, 8].

Die Probleme der Semantischen Suche sind denen der Sozialen Suche ähnlich. Zum einen ist das Erstellen von zusätzlichen Beschreibungen sehr aufwändig sowie fehleranfällig und zum anderen fehlen im WWW entsprechende Kontrollen. Außerdem ist das Modellieren von den Beziehungen der Begriffe so komplex, dass eine manuelle Modellierung nur durch Personen mit entsprechenden Kenntnissen durchgeführt werden kann und eine automatische Modellierung ausgereifte Methoden erfordert. Sowohl die manuelle als auch die automatische Modellierung sind noch nicht weit genug entwickelt, so dass semantische Suchansätze bisher nur für einzelne Dokumentensammlungen⁸⁶ oder in vereinfachter Form (z. B. durch die Anzeige verwandter Begriffe) angeboten werden.

3.1.8 Visuelle Suche

Visuelle Suchansätze versuchen die Suchqualität durch graphische Darstellung von Informationen zu verbessern [WBHÖ+09, 252]. Visualisierungen werden sowohl zur Unterstützung von Suchanfragen als auch bei der Darstellung der Suchergebnisse eingesetzt [WBHÖ+09, 249]. Die graphische Darstellung von Ergebnissen, Kontexten und weiterführenden Informationen kann dazu geeignet sein, dem Nutzer komplexe Zusammenhänge zu erklären und diesen bei der Optimierung seiner Suchanfrage zu unterstützen [WBHÖ+09, 251]. Beispiele für visuelle Suchansätze sind die Ergebnisvorschau (verkleinerte Anzeige der Webseiten), die Kategorisierung von Suchergebnissen (Clustering), die Darstellung von Relationen, die Visualisierung kollaborativer Komponenten (Tag-Clouds⁸⁷) sowie die Visualisierung zeitlicher und geografischer Aspekte [WBHÖ+09, 252 ff.].

Ob die Websuche durch Visualisierungen zu effizienter ermittelten Suchergebnissen oder einer Steigerung der Nutzerzufriedenheit gegenüber der traditionellen listenbasierten Ergebnisdarstellung führt, konnte allerdings noch nicht nachgewiesen werden [WBHÖ+09, 252].

⁸⁶ Der Websuchdienst Powerset beschränkt die Semantische Suche z. B. auf Wikipedia Artikel. <http://www.powerset.com>, Abruf: 2010-07-30

⁸⁷ „A Tag-Cloud is a list of the most popular tags, usually displayed in alphabetical order, and visually weighted by font size“ [HaHe06, o. S.].

3.1.9 Personalisierte Suche

Als Personalisierte Suche wird die Einbeziehung von Nutzerpräferenzen in die Suche beschrieben. Nach Griesbaum geht es bei der Personalisierten Suche darum, „individuelle Vorlieben und Bedürfnisse zu erkennen und bei der Ausgabe von Suchergebnissen zu berücksichtigen“ [Grie07, 13]. Die Erkennung von Nutzerpräferenzen kann durch impliziter (automatisierte Beobachtung) oder expliziter (Konfigurieren des Websuchdienstes durch den Nutzer) Gewinnung von Daten erfolgen [RiBr09, 154]. Die Anpassung des Suchergebnisses kann sich in einer reduzierten Ergebnismenge und einer veränderten Ergebnisreihenfolge oder -präsentation widerspiegeln.

Aus der Personalisierung der Suche können sowohl Nutzer als auch Anbieter von Websuchdiensten einen Vorteil ziehen. Die Nutzer erhalten individuelle, an ihre Bedürfnisse angepasste, Suchfunktionen und -ergebnisse. Die Anbieter können von einer verbesserten Vermarktung von Werbeanzeigen profitieren, da diese, durch die Verwertung von Nutzerdaten, an die Interessen des jeweiligen Nutzers angepasst werden können [Karo07, 155 f.].

Probleme der Personalisierten Suche stellen zum einen datenschutzrechtliche Bedenken auf Grund des Umgangs mit sensiblen Daten dar [GBRi09, 39] und zum anderen die starke Bindung des Nutzers an bestimmte Websuchdienste (Lock-In-Effekt)⁸⁸.

3.1.10 Lokale Suche

Bei der Lokalen Suche wird die Beantwortung einer Suchanfrage in Verbindung mit geografischen Daten durchgeführt. Die Lokale Suche dient sowohl Informationssuchenden als auch Anbietern von Produkten oder Dienstleistungen⁸⁹ und kann u. a. zum Auffinden von Adressen, Anbietern von Produkten oder Dienstleistungen oder zur Routenplanung genutzt werden [Lewa09c, 62]. Als Ergebnis werden dem Nutzer umgebungsangepasste Informationen angezeigt. Dazu wird die Umgebung entweder automatisch (z. B. über die Internetprotokoll-Adresse eines stationären Gerätes oder durch die Lokalisierung eines mobilen Gerätes) bestimmt oder durch den Nutzer angegeben [GBRi09, 38]. Die Ergebnisse werden durch eine Kombination von Daten aus Karten, Branchenbüchern und

⁸⁸ Ein Nutzer kann an einen Websuchdienst gebunden sein, wenn seine Nutzerdaten nur für die Anpassung des jeweiligen Websuchdienstes verwertet werden [Grie07, 25].

⁸⁹ Vgl. <http://searchenginewatch.com/3634952> und <http://searchenginewatch.com/3635027>, Abruf: 2010-07-23

Bewertungsportalen dargestellt.⁹⁰ Bei der Bewertung der Ergebnisse durch den Suchdienst kann es problematisch sein, zu bestimmen, ob für den Nutzer die Entfernung oder andere Kriterien des gesuchten Produktes bzw. der gesuchten Dienstleistung wichtiger sind. Häufig wird die Lokale Suche mit der Mobilien Suche⁹¹ verbunden [TKVK+08, o. S.].

3.1.11 Schlussfolgerungen

In diesem Abschnitt wurden verschiedene Suchansätze vorgestellt, die in den Konzepten von Websuchdiensten unterschiedlich stark verwendet und kombiniert werden können.

Wie sich gezeigt hat, ist eine allgemeingültige Abgrenzung von Websuchdiensten auf Grund der verschiedenen Ebenen, auf denen Websuchdienste miteinander verglichen werden können, nicht möglich.

3.2 Kategorien

In diesem Abschnitt werden Kategorien von alternativen Websuchdiensten benannt und die Eigenschaften dieser Kategorien erläutert.

Eine allgemeingültige Kategorisierung von alternativen Websuchdiensten existiert nicht. Dieses kann durch die Ausführungen aus dem vorherigen Abschnitt erklärt werden. Zudem ist zu beachten, dass die Entwicklung von Websuchdienstkategorien einer hohen Dynamik unterliegt. Deshalb können in kurzer Zeit neue Kategorien entstehen, welche auf neuen, weiterentwickelten oder kombinierten Suchansätzen basieren. Eine vollständige Darstellung der Kategorien alternativer Websuchdienste kann daher nur angestrebt werden. Damit der Vergleich alternativer Websuchdienste innerhalb von Kategorien durchgeführt werden kann, werden im Folgenden Kategorien dargestellt, die sich an den von Lewandowski [Lewa09c, 59 ff.], Griesbaum, Bekavac und Rittberger [GBRi09, 18 ff.; Beka09, o. S.] untersuchten Kategorien alternativer Websuchdienste orientieren.

Die Kategorien allgemeine Suchmaschinen und allgemeine Verzeichnisdienste werden nicht zu den Kategorien alternativer Websuchdienste gezählt. Die Eigenschaften und Funktionsweisen dieser beiden bekannten Websuchdienstkategorien wurden in der Literatur bereits mehrfach und ausführlich erläutert [BaRi99, 373 ff.; Ferb03, 299 ff.;

⁹⁰ Vgl. zum folgenden Abschnitt [Lewa09c, 62].

⁹¹ Der Begriff Mobile Suche umschreibt die Informationssuche im WWW über ein mobiles Endgerät.

Lewa09c, 54 ff.]. Die Entwicklung der allgemeinen Suchmaschinen wird jedoch zunehmend durch die Erweiterung um alternative Suchansätze geprägt (vgl. Abschnitt 2.3.4). Ebenso werden häufig Suchergebnisse aus alternativen Websuchdiensten oder speziellen Datenbanken in das Ergebnis der Standardsuche integriert. Damit soll dem Nutzer die Wahl der Suchfunktion abgenommen werden, die für die Suchanfrage das beste Ergebnis liefert, und das Problem der Tab Blindness⁹² umgangen werden [GBRi09, 43]. Deshalb können die allgemeinen Suchmaschinen der großen Anbieter auch als Universalsuchdienste bezeichnet werden [GBRi09, 46].

Im Gegensatz dazu versuchen alternative Websuchdienste, gezielt spezielle Dokumenttypen, bestimmte Bereiche des WWW oder speziell strukturierte Daten zu erschließen. Dazu wenden diese verschiedene Verfahren an, um mit den Vor- und Nachteilen, die sich aus den Eigenschaften des WWW, den Nutzern von Websuchdiensten (vgl. Abschnitt 2.2) und den Entwicklungen des Web 2.0 ergeben können (vgl. Abschnitt 3.1.5), umzugehen.

3.2.1 Bilder-, Video- und Audiosuchdienste

Bekanntere Kategorien von alternativen Websuchdiensten stellen die Multimediasuchdienste dar. Dazu gehören die Bilder-, die Video- und die Audiosuchdienste. Während die Bilder- und die Videosuchdienste von den meisten allgemeinen Suchmaschinen mit angeboten werden, sind die Audiosuchdienste nur noch vereinzelt bei allgemeinen Suchmaschinen zu finden [Lewa09c, 66]. Stattdessen existiert im WWW ein großes Angebot an kommerziellen Musikdatenbanken, über die urheberrechtlich geschützte Audiodateien vertrieben werden.⁹³ Allen drei Kategorien ist gemein, dass sich die Suche auf Dokumente bezieht, die für die Erschließung keine oder nur eingeschränkte textuelle Informationen bieten [Lewa09c, 66]. Deshalb wird deren Erschließung im WWW überwiegend mit Hilfe der Umgebungstexte durchgeführt [LeHö08, 2]. Dazu werden z. B. die Texte der Webseite auf der sich das Dokument befindet, die HTML Tags und die URL zur Auswertung herangezogen [LeHö08, 2 f.].

⁹² Tab Blindness beschreibt die Tatsache, dass eine große Anzahl der Nutzer von Suchdiensten, die zusätzlich angebotenen Suchfunktionen nicht wahrnehmen oder verstehen. Vgl. <http://searchenginewatch.com/3115131>, Abruf: 2010-08-28

⁹³ Siehe z. B. <http://new.music.yahoo.com/>, Abruf: 2010-09-10

Außerdem existieren für Bilder und Videos im WWW sehr große und populäre Communities, Plattformen und Archive. Deren Nutzer erstellen eine Vielzahl von Tags, die von Websuchdiensten verwertet werden können. Ebenso wird über diese Communities, Plattformen und Archive zum Teil eine so große Menge an Inhalten zur Verfügung gestellt, dass die Suche nach Bildern oder Videos über die interne Suchfunktion der Community, Plattform oder des Archivs ebenfalls eine Vielzahl von relevanten Ergebnissen liefern kann.

3.2.2 Personensuchdienste

Personensuchdienste unterstützen den Nutzer bei der Suche nach Informationen über Personen. Dazu aggregieren Personensuchdienste unterschiedliche Quellen. Neben den Daten, die auch über andere Websuchdienste, wie den allgemeinen Suchmaschinen, erfasst werden können, sammeln die Personensuchdienste speziell Daten aus sozialen Netzwerken, Blog- und Twitterdiensten, Nachrichtendiensten, Telefonbüchern und weiteren Datenbanken, die dem Deep Web zugerechnet werden können [Karz10, o. S.]. Die Ergebnisse werden strukturiert zusammengefasst und können z. B. Namen, Adresse, E-Mail-Adresse, Telefonnummer, Bilder und Interessen von Personen sowie deren Beziehungen zu weiteren Personen enthalten.⁹⁴ Da mehrere Personen einer Suchanfrage entsprechen können, ist es die Aufgabe des Websuchdienstes den Nutzer bei der Auswahl der gesuchten Person zu unterstützen [KNMe08, 1].

3.2.3 Frage-Antwort-Dienste

In Frage-Antwort-Diensten können individuelle Fragen gestellt oder bereits beantwortete Fragen durchsucht werden [SkMa08, 7]. Frage-Antwort-Dienste bauen besonders stark auf neu generierten Wissensbeständen auf.⁹⁵ Die Antworten auf neue Fragen basieren auf Informationen, die durch eine Gemeinschaft von Nutzern angegeben werden. Zwischen dem Stellen einer Frage und deren Beantwortung durch andere Nutzer vergeht im Vergleich zu anderen Websuchdiensten eine gewisse Zeit.

Der Vorteil von Frage-Antwort-Diensten besteht darin, dass die Nutzer ihre Suchanfragen natürlich sprachlich und ausführlicher umschreiben können als dies bei allgemeinen

⁹⁴ Siehe z. B. <http://www.123people.de>, Abruf 2010-09-01

⁹⁵ Vgl. zum folgenden Abschnitt [Lewa08, 2].

Suchmaschinen möglich ist. Zudem kann gezielt nach Meinungen und Ratschlägen in Bezug auf das individuelle Informationsbedürfnis gefragt werden [MaLe09, 53].

Durch die Beantwortung von Fragen im WWW können sich einzelne Personen als Experten zu bestimmten Themen herausstellen. In Kombination mit Ansätzen aus der Personensuche lässt sich daraus eine Expertensuche ableiten [Beka09, o. S.].

3.2.4 Social Bookmarking-Dienste

Social Bookmarking-Dienste sind primär für die Speicherung und Verwaltung von Lesezeichen⁹⁶ entwickelt worden [Lewa08, 2]. Durch die Verschlagwortung und Veröffentlichung der eigenen Lesezeichen können allerdings auch verschiedene Suchfunktionen angeboten werden.⁹⁷ Zum einen kann durch die Eingabe von Schlagwörtern in ein Suchfeld nach Lesezeichen gesucht werden. In dem daraufhin angezeigten Suchergebnis, wird die Reihenfolge der angezeigten Ergebnisse u. a. dadurch bestimmt, wie oft ein Lesezeichen durch unterschiedliche Nutzer gespeichert wurde. Somit werden die populärsten Inhalte zum Suchergebnis zuerst angezeigt werden. Zum anderen können populäre Tags, die in einer Tag-Cloud dargestellt werden, sowie Listen von populären und neuen Lesezeichen, als Suchvorschläge genutzt werden [GBRi09, 25].

Die Probleme der Social Bookmarking-Dienste ergeben sich aus einer unkontrollierten und subjektiven Verschlagwortung der Lesezeichen durch die Nutzer (vgl. Abschnitt 3.1.5).

3.2.5 Blog- und Twittersuchdienste

Blogsuchdienste können genutzt werden, um aktuelle Meinungen oder Kommentare von Nutzern zu aktuellen Ereignissen zu durchsuchen. Da es sich dabei in der Regel um zeitkritische Informationen handelt, sollte deren Erfassung und Bereitstellung zeitnah erfolgen.⁹⁸ Dazu werden neue Inhalte von Blogs nicht durch die Verfahren der Suchdienste ermittelt, sondern die Blogsoftware meldet das Einstellen eines neuen Beitrages automatisch an die Suchdienste. Ein Teil der Suchdienste kann Blogs auch automatisch

⁹⁶ Mit einem Lesezeichen (Bookmark) speichert der Nutzer die Adresse einer Webseite, um diese zu archivieren oder schneller abrufen zu können.

⁹⁷ Vgl. zum folgenden Abschnitt [SkMa08, 6].

⁹⁸ Vgl. zum folgenden Abschnitt [Lewa09c, 57; Lewa09c, 60 f.].

identifizieren. In diesem Fall werden die Blogs mit einem zusätzlichen Crawler in regelmäßigen Abständen nach Veränderungen durchsucht [ThHa07, 3].

Die Inhalte von Blogs können eine sehr heterogene Qualität aufweisen, da es jedem Nutzer möglich ist, einen Blog zu erstellen und zu führen. Die Strukturen von Blogs sind durch die Blogsoftware vorgeschrieben, so dass sich bestimmte Informationen einheitlich entnehmen und auswerten lassen. Dazu gehören z. B. Überschriften, Autorennamen und Datum der Blogbeiträge.

Neben der Blogsuche nach bestimmten Kriterien, z. B. nach Datumsangaben, können Blogsuchdienste auch zur Untersuchung von Trends genutzt werden.⁹⁹ Durch die Auswertung der Verlinkungen zwischen Blogs und der darin enthaltenen Kommentare kann ein Suchdienst die Popularität von Blogs ermitteln. Zudem kann durch die Aggregation der einzelnen Blogs eine öffentliche Meinung abgebildet werden. Unternehmen können Blogsuchdienste z. B. dazu nutzen, um ein schnelles Feedback über ihre Produkte zu bekommen. Allerdings handelt es sich um Verlinkungen und Meinungen von Blog-affinen Nutzern, die ausgewertet werden. Zudem können die Einträge auch durch finanzielle Interessen subjektiv beeinflusst sein.

Neben Blogs existiert eine weitere Möglichkeit, mit der die Nutzer ihre Meinungen und Kommentare im WWW publizieren können. Über den Webdienst Twitter können, nach der Erstellung eines Profils, Twitter-Nachrichten (Tweets) verfasst und veröffentlicht werden. Twitter-Nachrichten unterscheiden sich von den Beiträgen in Blogs durch ihre stark begrenzte Anzahl von Zeichen. Die Erschließung und der Funktionsumfang von Twittersuchdiensten sind denen der Blogsuchdienste sehr ähnlich.

3.2.6 Nachrichtensuchdienste

Nachrichtensuchdienste werden ebenfalls zum Auffinden von zeitkritischen Informationen genutzt.¹⁰⁰ Im Unterschied zu den Blogs und Tweets handelt es sich bei Nachrichten um Dokumente, die durch ihre journalistische Erstellung neben der strukturellen auch eine gewisse inhaltliche Qualität aufweisen. Dadurch können ebenso die Inhalte der Nachrichten durch einheitliche Verfahren erschlossen werden. Für die Erschließung der Nachrichten wird ein eigener Crawler angelegt, der nur ausgewählte Quellen im WWW auf

⁹⁹ Vgl. zum folgenden Abschnitt [The109, 48 ff.].

¹⁰⁰ Vgl. zum folgenden Abschnitt [Lewa09c, 59 f.].

neue Meldungen durchsucht. Durch die eingegrenzte Anzahl von Webseiten werden diese Quellen schneller und in kürzeren Abständen durchsucht als dies durch Crawler für das gesamte WWW möglich ist.¹⁰¹ Die Ergebnisse von Nachrichtensuchdiensten können neben den erschlossenen Nachrichten auch zugekaufte Ergebnisse enthalten, die durch den eigenen Crawler nicht auffindbar sind. Ob die Nachrichtensuchdienste allerdings eine vollständige Abdeckung der Nachrichten im WWW bieten können, lässt sich nicht überprüfen, da die Anbieter der Nachrichtensuchdienste häufig keine öffentlichen Quellenlisten anbieten.

3.2.7 Wissenschaftssuchdienste

Wissenschaftssuchdienste verfolgen das Ziel, Dokumente mit wissenschaftlichen Inhalten auffindbar zu machen [Lewa09c, 62 f.]. Da der Zugriff auf diese Dokumente kosten- oder registrierungspflichtig sein kann oder diese in Datenbanken verwaltet werden und darüber abgerufen werden müssen, kann ein Teil der wissenschaftlichen Dokumente dem Deep Web zugerechnet werden [LeMa06, 3 ff.]. Die Verfahren der allgemeinen Suchmaschinen können diese Dokumente auf Grund technischer Hindernisse nicht erfassen. Wissenschaftssuchdienste versuchen über einen vordefinierten Suchraum und in Abstimmung mit den Datenbanken wissenschaftlicher Dokumente, diese Dokumente, oder zumindest Teile der Dokumente, durchsuchbar zu machen (vgl. Abschnitt 3.1.4). Handelt es sich bei den Ergebnissen um kostenpflichtige Inhalte, werden dem Nutzer Auszüge oder Kurzbeschreibungen der Inhalte angezeigt sowie Möglichkeiten zum Erwerb der vollständigen Inhalte.

3.2.8 Schlussfolgerungen

In diesem Kapitel konnte ein Überblick über Kategorien alternativer Websuchdienste gewonnen werden. Die Ausführungen haben verdeutlicht, dass alternative Websuchdienste auf sehr unterschiedliche Nutzerbedürfnisse, Suchobjekte und Bereiche des WWW ausgerichtet sind und dazu verschiedene Suchtechniken verwenden. Ein Vergleich von Websuchdiensten sollte deshalb innerhalb von Websuchdienstkategorien erfolgen.

¹⁰¹ Mit allgemeinen Crawlern wird versucht, alle Webseiten und Inhalte im WWW zu erschließen.

4 Vergleich alternativer Websuchdienste

4.1 Ermittlung der Grundgesamtheit

In dieser Arbeit wird ein Vergleich von alternativen Websuchdiensten innerhalb der folgenden zehn Websuchdienstkategorien durchgeführt: Bildersuchdienste, Videosuchdienste, Audiosuchdienste, Personensuchdienste, Frage-Antwort-Dienste, Social Bookmarking-Dienste, Blogsuchdienste, Twittersuchdienste, Nachrichtensuchdienste und Wissenschaftssuchdienste. Für den Vergleich wird zuerst eine Grundgesamtheit an alternativen Websuchdiensten, die diesen ausgewählten Kategorien entsprechen, ermittelt. In diesem Abschnitt wird die Ermittlung der Grundgesamtheit beschrieben.

Als erstes werden Anforderungen an die Grundgesamtheit gestellt, die bei der Wahl der Methodik zur Ermittlung der Grundgesamtheit zu berücksichtigen sind. Die Grundgesamtheit an alternativen Websuchdiensten soll möglichst vollständig und aktuell sein sowie objektiv ermittelt werden.

Vollständigkeit bedeutet, dass die Grundgesamtheit möglichst alle im WWW verfügbaren Suchdienste, die einer der Websuchdienstkategorien zugeordnet werden können, umfasst. Durch diese Anforderung soll jedem Suchdienst, der einer ausgewählten Kategorie entspricht, die Möglichkeit gegeben werden, in das Ranking aufgenommen zu werden.

Weiter muss in der Grundgesamtheit das aktuellste Angebot alternativer Websuchdienste abgebildet sein. Alternative Websuchdienste unterliegen, ebenso wie die Dokumente im WWW, einer hohen Dynamik. Das Angebot an alternativen Websuchdiensten kann sich häufig und in kurzer Zeit ändern, da zum einen die Nachfrage nach bestimmten Arten von Inhalten sich verändern kann und zum anderen die Kosten für den Betrieb eines alternativen Websuchdienstes nur bei ausreichender Attraktivität für den Werbemarkt durch Werbeeinnahmen gedeckt werden können (vgl. Abschnitt 2.3.3). Auf Grund der bisher geringen Nutzung von alternativen Websuchdiensten [GBRi09, 43] und der daraus folgenden geringen Attraktivität für den Werbemarkt kann es vorkommen, dass alternative Websuchdienste nach kurzer Zeit wieder aus dem Markt für Websuchdienste ausscheiden. Für einen aktuellen Vergleich von alternativen Websuchdiensten ist es deshalb wichtig, eine aktuelle Grundgesamtheit an Websuchdiensten zu verwenden.

Schließlich darf die Grundgesamtheit keiner subjektiven Auswahl an alternativen Websuchdienste entsprechen. Finanzielle Interessen und subjektive Beurteilungen von einzelnen Personen oder kleinen Gruppen von Personen sollen, soweit es möglich ist, ausgeschlossen werden, um eine neutrale Grundgesamtheit für den Vergleich zu erhalten.

In dieser Arbeit wird die Grundgesamtheit an alternativen Websuchdiensten durch zwei Verfahren ermittelt. Zum einen werden Websuchdienstlisten von fachbezogenen Quellen¹⁰² im WWW herangezogen (vgl. Abschnitt 4.1.1) und zum anderen werden Suchanfragen für jede ausgewählte Kategorie an die allgemeine Suchmaschine Google gestellt, um die dazugehörigen Suchergebnisse auswerten zu können (vgl. Abschnitt 4.1.2). In wie weit diese beiden Verfahren für die Erfüllung der aufgestellten Anforderungen an die Grundgesamtheit geeignet sind, wird im Folgenden erläutert.

Vollständigkeit kann weder durch die Websuchdienstlisten noch durch die Suchanfragen an allgemeine Suchmaschinen erreicht werden. Die Websuchdienstlisten enthalten zwar eine große Anzahl von alternativen Websuchdiensten, leisten jedoch keine vollständige Darstellung aller verfügbaren alternativen Websuchdienste.¹⁰³ Allgemeine Suchmaschinen können zwar ebenfalls keine vollständige Abdeckung des WWW erreichen, jedoch besitzen diese die höchste Abdeckung unter den Websuchdiensten [GBRi09, 26 ff.]. Zudem sollen in dieser Arbeit die populärsten alternativen Websuchdienste dargestellt werden. Die angezeigte Ergebnisreihenfolge der allgemeinen Suchmaschine Google berücksichtigt in hohem Maße die Popularität der Ergebnisse [LaMe06, 13] und stellt deshalb ein geeignetes Instrument zur vollständigen Ermittlung der populärsten alternativen Websuchdienste dar. Auch wenn keine vollständige Erfassung aller alternativen Websuchdienste möglich ist, so ist trotzdem zu erwarten, dass durch die Anwendung dieser beiden Verfahren eine große Anzahl von alternativen Websuchdiensten ermittelt werden kann und die populärsten Websuchdienste in dieser Arbeit erfasst werden.

Die Aktualität der Websuchdienstlisten kann nur zum Teil festgestellt werden, da nur zwei Quellen den Zeitpunkt der Erstellung der Websuchdienstliste angeben. Die Aktualität der enthaltenen Websuchdienste kann jedoch geprüft werden, indem die Erreichbarkeit der

¹⁰² http://sprint.informationswissenschaft.ch/fileadmin/sprint/texte/Suchmaschinen_im_WWW.pdf,
<http://www.philb.com/whichengine.htm>, <http://www.pandia.com/powersearch/index.html>,
http://www.readwriteweb.com/archives/top_100_alternative_search_engines_mar07.php,
<http://websearch.about.com/od/enginesanddirectories/tp/search-engine-list.htm>,
<http://www.fasteagle.com/search/index.php#> und <http://www.panabee.com/>, Abruf: 2010-09-06

¹⁰³ Die Autoren der Websuchdienstlisten bieten nur eine Auswahl an Websuchdiensten an.

Websuchdienste kontrolliert wird. Durch Suchanfragen an allgemeine Suchmaschinen können neue Websuchdienste gefunden werden, da der Index von allgemeinen Suchmaschinen in regelmäßigen Abständen durch automatische Verfahren aktualisiert wird.¹⁰⁴ Die Suchmaschine von Google verfügt über einen sehr großen sowie aktuellen Index und wird in dieser Arbeit für die Ermittlung einer aktuellen Grundgesamtheit verwendet.¹⁰⁵

Die Objektivität der einzelnen Websuchdienstlisten kann nicht nachgewiesen werden. Allerdings stammen die Websuchdienstlisten aus sieben verschiedenen und voneinander unabhängigen Quellen. Deshalb relativiert sich die Subjektivität einer einzelnen Quelle durch die Zusammenführung dieser sieben Quellen. Die Objektivität der Suchmaschinen-ergebnisse lässt sich ebenfalls nicht überprüfen, da die konkreten Verfahren der Suchmaschinen von den Betreibern zur Wahrung des Geschäftsgeheimnisses nicht veröffentlicht werden. Bekannt ist jedoch, nach welchen Prinzipien die Verfahren der Suchmaschinen bei der Beantwortung der Suchanfrage vorgehen. Während die Ergebnismenge abhängig von der Suchanfrage ist, ist die Bewertung der Ergebnisreihenfolge unabhängig von der Suchanfrage und erfolgt nach algorithmischen Verfahren [Lewa05, 90 ff.]. Diese algorithmischen Verfahren werten zwar eine Vielzahl von subjektiven Beurteilungen aus, jedoch stammen diese subjektiven Beurteilungen nicht von nur einer einzelnen Person oder einer kleinen Gruppe von Personen. Zudem werden in der Ergebnisanzeige von Google organische Ergebnisse¹⁰⁶ und gesponserte Ergebnisse (bezahlte Werbeanzeigen) deutlich voneinander getrennt. Somit lassen sich subjektive Einflüsse auf die Grundgesamtheit zwar nicht ausschließen, jedoch wird der Einfluss einer einzelnen Quelle durch die Anwendung beider Verfahren und deren Zusammenführung zu einer Grundgesamtheit verringert.

In den folgenden zwei Abschnitten werden die Auswertung der Websuchdienstlisten und die Ermittlung von alternativen Websuchdiensten über Suchanfragen an die allgemeine Suchmaschine beschrieben.

¹⁰⁴ Vgl. zum folgenden Abschnitt [Lewa04a, 310 ff.]

¹⁰⁵ Die allgemeinen Suchmaschinen von Bing und Yahoo verfügen ebenfalls über einen großen und aktuellen Index. Die Suchergebnisse (nach dem Verfahren in Abschnitt 4.1.2) dieser beiden Suchmaschinen enthielten allerdings eine zu geringe Anzahl von alternativen Websuchdiensten, so dass diese nicht zur Ermittlung der Grundgesamtheit hinzugezogen wurden.

¹⁰⁶ Die organischen Ergebnisse werden nach den algorithmischen Verfahren der Suchmaschine ermittelt.

4.1.1 Websuchdienstlisten

Die Websuchdienstlisten wurden aus fachbezogenen Quellen im WWW entnommen und enthalten Namen und/oder URL von Websuchdiensten sowie eine von den Autoren zugeordnete Kategorie. Abb. 5 zeigt als Beispiel einen Ausschnitt aus der Websuchdienstliste von Pandia Powersearch.

| Images and media | TV and video | Music & MP3 |
|---|--|--|
| AlltheWeb Pictures Alta Vista Images Bing Images Compfight image search Creative Commons on Flickr FindSounds.com Flickrstorm Google Audio Indexing Google Images Google Image Ripper Mlticolor Search Lab (visual Flickr search) picsearch Pixsy Speedcine free movies Real Guide TinEye (reverse image search) NEW! V like Vintage (historic photos) Webseek Yahoo! Audio Search Yahoo Image Search more... | AOL Video Bing (Live) Video Blinkx TV Blip TV ClipBlast ClipShack Clipta Dabble Flixya Google Video Guba Hulu (US only) meefedia meevee Mega Video Metacafe MySpace Video OVGuide Putpile Search for Video Singing Fish Stickam trooker NEW! TV Bookmark uvouch Veoh Videosurf Vidiac Yahoo! Video YouTube more... | <p>P = fee based</p> Apple iTunes (requires download) P Audiogalaxy P Bloson eMusic P Magnatune Mp3.com music-map (for related artists) Napster P SkreemR Songza NEW! Spotify (requires download) P Y! Music more mp3... |
| Radio Pandia Radio Search BRS Radio Directory Live Radio Virtual Tuner Radio Locator (MIT) vTuner more... | | Software w=Win m=Mac *=other cnet download.com wm* Freeware Planet Jumbo w Tucows wm* ZDNet w more... |
| Podcasting podanza Podcast directory Podcast network Podscope | | |

Abb. 5: Ausschnitt aus der Websuchdienstliste von Pandia Powersearch¹⁰⁷

Die Anzahl der in den sieben Websuchdienstlisten enthaltenen Websuchdienste beträgt zum Zeitpunkt, an dem mit der Auswertung begonnen wurde (20. September 2010), insgesamt 1695 Websuchdienste.¹⁰⁸ Um eine Auswertung der Websuchdienstlisten durchführen zu können, werden zuerst die Bezeichnungen für die Kategorien normiert. Die Kategorien werden entsprechend der Kategorien aus Abschnitt 3.2 normiert, da in den sieben

¹⁰⁷ <http://www.pandia.com/powersearch/index.html>, Abruf: 2010-09-06

¹⁰⁸ Einige Websuchdienste sind in mehr als einer Quellen enthalten. Deshalb kann die genannte Gesamtanzahl Dubletten enthalten.

voneinander unabhängigen Quellen keine einheitliche Bezeichnung für die Kategorien von den Autoren verwendet wurde.¹⁰⁹

Danach werden Kriterien aufgestellt, an denen geprüft wird, ob ein aufgelisteter Websuchdienst in die Grundgesamtheit für den Vergleich alternativer Websuchdienste aufgenommen oder davon ausgeschlossen wird. Diese Kriterien sind zum einen allgemeine Kriterien, die für alle untersuchten Websuchdienstkategorien gelten und zum anderen kategorieabhängige Kriterien, die für einzelne Kategorien gelten. Die allgemeinen Kriterien ergeben sich aus den folgenden Ausführungen.

Als erstes müssen die aufgelisteten Websuchdienste aktuell erreichbar und funktionsfähig sein, um für den Vergleich übernommen zu werden. Weiter müssen sie den Beschreibungen der jeweiligen Kategorie aus Abschnitt 3.2 entsprechen.

Ausgeschlossen werden Websuchdienste, die nur die Suchergebnisse anderer Websuchdienste anzeigen.¹¹⁰ Zu unterscheiden von der Nutzung der Suchergebnisse anderer Websuchdienste ist die Nutzung der Indizes anderer Websuchdienste. Ein Websuchdienst kann den Index eines anderen Websuchdienst nutzen, um diesen nach eigenen Kriterien auszuwerten und die Ergebnisse mit eigenen Mitteln darzustellen. Diese Websuchdienste erbringen eine ausreichend eigenständige Leistung, um für den Vergleich übernommen zu werden. Zu beachten ist allerdings, dass die Anzeige externer Suchergebnisse sowie die Nutzung externer Indizes nicht immer von den Websuchdiensten kenntlich gemacht wird. Deshalb können bei der Prüfung dieses Kriteriums nur Websuchdienste ausgeschlossen werden, bei denen die Anzeige externer Suchergebnisse erkennbar ist.

Ebenso werden Websuchdienste vom Vergleich ausgeschlossen, die nur Inhalte mit Bezug auf ein bestimmtes Land oder ein bestimmtes Thema durchsuchbar machen.¹¹¹ Zwar ist die Eigenschaft „Spezialisierung“ ein Merkmal alternativer Websuchdienste, jedoch wurde diese Spezialisierung bereits durch die Kategorisierung in Abschnitt 3.2 hervorgehoben und wird in dieser Untersuchung nicht weiter unterteilt.

¹⁰⁹ Die von den Autoren verwendeten Kategorien academic, scientific und scholarly wurden z. B. unter der Kategorie Wissenschaftssuchdienste zusammengefasst. Hingegen wurden Websuchdienste, die von den Autoren in einer Kategorie zusammengefasst sind (z. B. Multimediasuchdienste), den einzelnen Kategorien (Bilder-, Video- und Audiosuchdienste) zugeordnet.

¹¹⁰ Es genügt z. B. nicht, wenn ein Websuchdienst nur eine eigene Suchoberfläche bereitstellt oder die Suchergebnisse mehrerer Websuchdienste auf einer Seite anzeigt.

¹¹¹ Es werden z. B. keine Personensuchdienste übernommen, die sich nur auf den US-amerikanischen Raum beziehen und keine Wissenschaftssuchdienste, die sich nur auf den Fachbereich Medizin beschränken.

Schließlich werden keine Websuchdienste aus den Websuchdienstlisten übernommen, bei denen die Nutzung der Suchfunktion kosten- oder registrierungspflichtig ist. Gleiches gilt für Websuchdienste, die überwiegend Inhalte im Suchergebnis anzeigen, deren Zugriff über den Websuchdienst kosten- oder registrierungspflichtig ist.¹¹² Einzige Ausnahme von diesem Kriterium stellen die Wissenschaftssuchdienste dar. Die Suchergebnisse von Wissenschaftssuchdiensten enthalten häufig Auszüge oder Kurzfassungen von Artikeln oder Büchern, deren vollständiger Inhalt kostenpflichtig sein kann.

Neben den genannten allgemeinen Kriterien werden folgende kategorieabhängige Kriterien aufgestellt.

In der Kategorie Bildersuchdienste werden auch Communities, Plattformen und Archive für Bilder übernommen, die eine Suchfunktion anbieten. Communities, Plattformen und Archive dienen neben der Speicherung von Dokumenten im WWW auch dazu, diese wieder auffindbar zu machen und anderen Nutzern zur Verfügung zu stellen. Gleiches gilt in der Kategorie Videosuchdienste für Communities, Plattformen und Archive für Videos.

In der Kategorie Audiosuchdienste werden diejenigen Websuchdienste übernommen, die in der Anzeige der Suchergebnisse eine kostenfreie Möglichkeit anbieten, mit der die angezeigten Audiodateien heruntergeladen werden können.¹¹³ Damit werden Dienste ausgeschlossen, die nur Informationen über Musik bereitstellen, zeitlich begrenzte Versionen zur Vorschau anbieten oder den Nutzer die Inhalte über einen Stream¹¹⁴ abspielen lassen.

Von der Kategorie Personensuchdienste werden soziale Netzwerke ausgeschlossen. Diese enthalten zwar eine Vielzahl von Informationen über Personen und bieten ebenfalls eigene Personensuchfunktionen an, jedoch stellt das Auffinden von Personen nur eine von vielen Funktionen von sozialen Netzwerken dar. Soziale Netzwerke würden den Vergleich von Personensuchdiensten verzerren, da die Popularität der Personensuchfunktion nicht getrennt von der großen Anzahl weiterer Funktionen von sozialen Netzwerken bestimmt werden kann. Ebenso werden keine auf das WWW übertragenen Telefon- oder

¹¹² Dies gilt z. B. für Bildersuchdienste, die nach geschützten Bildern suchen und diese zum Verkauf anbieten.

¹¹³ Inwieweit einzelne Audiosuchdienste gegen das Urheberrecht verstoßen, falls mit ihrer Hilfe urheberrechtlich geschützte Audiodateien kostenfrei heruntergeladen werden können, wird im Rahmen dieser Arbeit nicht kontrolliert und wurde deshalb nicht als ein Kriterium berücksichtigt.

¹¹⁴ Über einen Stream können Daten wiedergegeben werden, ohne die ganze Datei auf dem Rechner des Nutzers zu speichern.

Branchenbücher als Personensuchdienste übernommen. Stattdessen werden diejenigen Suchdienste als Personensuchdienste übernommen, die mehrere dieser Quellen¹¹⁵ nutzen, um die Suchergebnisse daraus zu aggregieren.

Für Nachrichtensuchdienste wurde ein ähnliches Kriterium aufgestellt. In die Grundgesamtheit werden keine Nachrichtendienste aufgenommen, die ausschließlich ihre eigene Website mit der eigenen Suchfunktion durchsuchbar machen. Wie bei den Personensuchdiensten werden also diejenigen Nachrichtensuchdienste übernommen, bei denen die Suche im Vordergrund steht und die Suchergebnisse aus mehreren Nachrichtendiensten aggregiert werden.

Durch die Auswertung der Websuchdienstlisten nach diesen allgemeinen und kategorieabhängigen Kriterien wurde der erste Teil der Grundgesamtheit für den angestrebten Vergleich bestimmt.

4.1.2 Suchanfragen an Google

Der zweite Teil der Grundgesamtheit wird durch die Formulierung von Suchanfragen an die allgemeine Suchmaschine Google und der daran anschließenden Auswertung der Suchergebnisse ermittelt. Zuerst ist auf die Formulierung der Suchanfragen einzugehen, da die Suchergebnisse von den Suchanfragen abhängig sind.

Ein Suchergebnis enthält Treffer, die mit den verwendeten Begriffen der Suchanfrage übereinstimmen.¹¹⁶ Deshalb müssen die Suchanfragen so formuliert sein, dass möglichst alle verfügbaren alternativen Websuchdienste der jeweiligen Kategorie gefunden werden (recall-orientierte Suchanfragen). Zum Auffinden geeigneter Begriffe, die für die Formulierung der Suchanfragen verwendet werden können, wird überprüft, mit welchen Begriffen sich die alternativen Websuchdienste selbst bezeichnen. Dazu werden zum einen die Webseiten¹¹⁷ von den alternativen Websuchdiensten untersucht, die in den Websuchdienstlisten (vgl. Abschnitt 4.1) enthalten sind, und zum anderen deren

¹¹⁵ Neben sozialen Netzwerken sowie Telefon- und Branchenbücher werden z. B. auch Blogs und Twitterprofile zur Aggregation von Daten verwendet.

¹¹⁶ Die Suchergebnisse, die einer Suchanfrage zugeordnet werden, werden durch verschiedene Verfahren von Suchdiensten ermittelt. Grundlage der Suchmaschinen ist allerdings das Boolesche Modell, auf das sich hier bezogen wird [Lewa05, 83].

¹¹⁷ Untersucht wurden die Hauptseite des Websuchdienstes und falls vorhanden, die Informationsseite „About“/„About Us“ des jeweiligen Websuchdienstes.

Suchergebniseintrag¹¹⁸ bei der allgemeinen Suchmaschine Google abgerufen. Alle dort enthaltenen Begriffe, die der Beschreibung von Websuchdiensten und deren Kategorien dienen, werden notiert und den entsprechenden Kategorien zugeordnet. Somit kann ein Überblick gewonnen werden, welche Begriffe in den Suchanfragen vorkommen müssen, damit möglichst viele alternative Websuchdienste gefunden werden können.

Nachdem für jede Kategorie die relevanten Begriffe festgelegt wurden, werden die Suchanfragen an die allgemeine Suchmaschine von Google¹¹⁹ gestellt. Verwendet wird der internationale Dienst¹²⁰ der Suchmaschine, da mit diesem die größte Anzahl von international ausgerichteten Websuchdiensten ermittelt werden kann.¹²¹ Die Suchbegriffe der jeweiligen Kategorie werden in der Suchanfrage durch den Suchoperator OR getrennt und zusammenhängende Wörter durch Anführungszeichen eingeschlossen (Phrasensuche).¹²² Durch die Verwendung des Suchoperators OR werden von der Suchmaschine alle indexierten Webseiten im Suchergebnis dargestellt, die mindestens mit einem der verwendeten Suchbegriffe übereinstimmen. In Abb. 6 wird als ein Beispiel für eine solche Suchanfrage ein Ausschnitt des Suchergebnisses für die Kategorie Bildersuchdienste dargestellt. Die verwendete Suchanfrage für die Bildersuchdienste lautet: "image search" OR "search images" OR "image search engine" OR "picture search" OR "search pictures" OR "picture search engine" OR "photo search" OR "search photos" OR "photo search engine".

Nach der Eingabe einer Suchanfrage werden die 100 ersten Suchergebnisse je Kategorie nach alternativen Websuchdiensten untersucht. Die Auswertung der Suchmaschinen-ergebnisse wird nach denselben allgemeinen und kategorieabhängigen Kriterien, wie sie in Abschnitt 4.1.1 genannt wurden, durchgeführt. Außerdem wird ein Websuchdienst nur

¹¹⁸ Als Suchergebniseintrag wird in dieser Arbeit der Titel und die Kurzbeschreibung bezeichnet, mit denen eine Webseite in dem Suchergebnis eines Suchdienstes erscheint.

¹¹⁹ <http://www.google.com/>, Abruf: 2010-09-07

¹²⁰ Internationale Dienste von Unternehmen, hier Anbieter alternativer Websuchdienste, sind im Allgemeinen unter der Top Level Domain com zu finden. Vgl. <http://tools.ietf.org/html/rfc1591>, Abruf: 2010-09-12

¹²¹ Die Suchergebnisse werden von den Suchmaschinen an die verwendete Top Level Domain der Suchmaschinen angepasst, indem die Ergebnisse, die zur verwendeten Sprache und der geografischen Umgebung der verwendeten Top Level Domain passen, im Suchergebnis bevorzugt werden [Lewa05, 70 ff.].

¹²² Vgl. zum folgenden Abschnitt <http://www.google.com/support/websearch/bin/answer.py?answer=136861>, Abruf: 2010-10-10

dann aus dem Suchergebnis übernommen, wenn dieser durch seinen eigenen Eintrag in dem Suchergebnis angegeben wird.¹²³



Abb. 6: Ausschnitt aus dem Suchergebnis zur Kategorie Bildersuchdienste

Nachdem durch diesen Vorgang die Suchergebnisse zu den formulierten Suchanfragen ausgewertet wurden, werden die Ergebnisse aus Abschnitt 4.1.1 und Abschnitt 4.1.2 zu einer Grundgesamtheit zusammengeführt und Dubletten innerhalb der Kategorien entfernt. Die auf diese Weise zusammengetragenen alternativen Websuchdienste stellen die Grundgesamtheit an alternativen Websuchdiensten für den Vergleich in dieser Arbeit dar.

Es konnte eine Grundgesamtheit von

- 50 Bildersuchdiensten,
- 45 Videosuchdiensten,
- 24 Audiosuchdiensten,
- 7 Personensuchdiensten,

¹²³ Ein Websuchdienst wird nicht übernommen, wenn dieser nur in den Beschreibungstexten von anderen Suchergebniseinträgen angezeigt wird.

- 13 Frage-Antwort-Diensten,
- 21 Social Bookmarking-Diensten,
- 17 Blogsuchdiensten,
- 27 Twittersuchdiensten,
- 25 Nachrichtensuchdiensten,
- 24 Wissenschaftssuchdiensten

ermittelt werden.

4.2 Kennzahlenerhebung

Nach Ermittlung der Grundgesamtheit werden für jeden darin enthaltenen Websuchdienst die Werte für die webometrischen Kennzahlen WIF, PageRank und Alexa Traffic Rank abgerufen.

Wie in Abschnitt 2.4 erörtert wurde, beziehen sich die Werte der Kennzahlen auf die Webseite, auf der ein Websuchdienst bereitgestellt wird. Die Adresse einer Webseite im WWW wird durch die URL einer Webseite angegeben und zum Abrufen der Kennzahlen benötigt. Dabei kann es vorkommen, dass ein Websuchdienst über unterschiedliche URLs erreichbar ist.¹²⁴ Die Werte der Kennzahlen sind wiederum abhängig von der angegebenen URL.¹²⁵ Wird ein Websuchdienst z. B. mit einer sehr komplexen URL angegeben, kann es vorkommen, dass die Werte für die Kennzahlen nicht abgerufen werden können oder nur ein sehr geringer Wert für den Websuchdienst ausgegeben wird. Da in den Websuchdienstlisten (vgl. Abschnitt 4.1.1) und in den Suchergebnissen (vgl. Abschnitt 4.1.2) die URLs der Websuchdienste in unterschiedlicher Form angegeben sind, werden die URLs von den Websuchdiensten der Grundgesamtheit vor dem Abruf der Kennzahlen in eine einheitliche Form gebracht (normalisiert).¹²⁶ Für die Normalisierung wird die URL eines Websuchdienstes gewählt, die beim Aufruf des Websuchdienstes in der Adresszeile des Browser angezeigt wird. Diese URL wird soweit gekürzt, dass weiterhin kein

¹²⁴ Die Bildersuche von Google ist z. B. über die URL <http://www.google.com/imghp> und über die URL <http://images.google.com/> erreichbar, Abruf 2010-10-09

¹²⁵ Für die URL <http://www.bing.com/images?FORM=Z9LH5> konnte z. B. nur ein WIF von 109 abgerufen werden, für die URL <http://www.bing.com/images> desselben Websuchdienstes hingegen ein WIF von 34325, Abruf: 2010-10-06

¹²⁶ Für weitere Ausführungen zur Normalisierung von URLs siehe z. B. [LKHo05, 1076 ff.].

zusätzlicher Navigationsschritt zum Erreichen des Websuchdienstes benötigt wird.¹²⁷ Falls eine internationale Version eines Websuchdienstes der Grundgesamtheit verfügbar ist, wird die URL der internationalen Version gewählt. Für diese normalisierten URLs der Websuchdienste werden die Werte der Kennzahlen abgerufen. Im Folgenden werden das Vorgehen und die verwendeten Instrumente für die drei Kennzahlen beschrieben.

4.2.1 Web Impact Factor

Der WIF eines Websuchdienstes wird in einer stark vereinfachten Form erhoben. Im Gegensatz zu dem WIF von Ingwersen (vgl. Abschnitt 2.4.1.1) wird in dieser Arbeit mit dem WIF nur die Anzahl externer Inlinks einer Webseite beschrieben. Aus der Sicht einer bestimmten Webseite werden als Inlinks alle Verlinkungen im WWW bezeichnet, die auf diese bestimmte Webseite verweisen [Björ04, 16]. Als extern werden alle Verlinkungen bezeichnet, die sich nicht auf der eigenen Website befinden. Durch diese Eingrenzung wird berücksichtigt, dass die Popularität der Websuchdienste eine Bewertung von außen darstellt.

Die Anzahl der externen Inlinks lässt sich mit Hilfe einer Suchmaschine bestimmen. Diese bieten Eingabebefehle oder Analysewerkzeuge an, die ein einfaches Abrufen der externen Inlinks ermöglichen. Dazu greifen die Suchmaschinen auf die Daten zurück, die durch den/die Crawler der Suchmaschine gesammelt wurden. Trotz der unvollständigen Abdeckung des WWW durch die Crawler (vgl. Abschnitt 2.2.1) stellen Suchmaschinen das beste zur Verfügung stehende Mittel für den Abruf der externen Inlinks dar [The104]. In dieser Arbeit wird dazu der Yahoo Site Explorer¹²⁸ verwendet. Für jeden Websuchdienst der Grundgesamtheit wird die URL des Websuchdienstes in das Eingabefeld des Yahoo Site Explorers eingegeben und die Ergebnisanzeige auf externe Inlinks beschränkt.¹²⁹ Daraufhin wird die angezeigte Anzahl externer Inlinks notiert.

¹²⁷ Die URL <http://www.bing.com/images?FORM=Z9LH5> konnte z. B. auf die URL <http://www.bing.com/images> gekürzt werden, Abruf: 2010-10-08.

¹²⁸ <http://siteexplorer.search.yahoo.com/>, Abruf: 2010-09-12

¹²⁹ Über die angezeigte Drop-down-Liste wurde die Option „Except from this domain“ ausgewählt.

4.2.2 PageRank

Der PageRank eines Websuchdienstes wird über die Google Toolbar¹³⁰ ermittelt. Die Google Toolbar ist eine Browsererweiterung und erstellt eine zusätzliche Symbolleiste im Browser. Über diese Symbolleiste werden von Google verschiedene Funktionen und Informationen bereitgestellt, mit denen die Anwender bei der Nutzung des WWW unterstützt werden sollen. In dieser Arbeit wird von der Google Toolbar der Wert des PageRanks einer Webseite abgelesen. Google gibt den PageRank einer Webseite als gerundeten ganzzahligen Wert an [LaMe06, 28]. Dieser liegt zwischen Null und Zehn. Zehn stellt den höchsten Wert dar und Null den kleinsten. Für jeden Websuchdienst der Grundgesamtheit wird die URL des Websuchdienstes über einen Browser mit installierter Google Toolbar aufgerufen und der daraufhin angezeigte Wert für den PageRank notiert.

4.2.3 Alexa Traffic Rank

Der Alexa Traffic Rank eines Websuchdienstes wird bei dem Webdienst von Alexa Internet¹³¹ abgerufen. Da der Wert des Alexa Traffic Rank den Rang einer Website in dem Ranking von Alexa Internet darstellt, ist der höchste Wert für den Alexa Traffic Rank der Rang eins. Der niedrigste Wert kann nicht bestimmt werden, da der letzte Rang des Rankings nicht abgelesen werden kann. Für jeden Websuchdienst der Grundgesamtheit wird die URL des Websuchdienstes in das Eingabefeld des Webdienstes von Alexa Internet eingeben und der daraufhin angezeigte Rang notiert. Der angezeigte Alexa Traffic Rank bezieht sich immer auf die Second Level Domain¹³² einer URL. Welche Bedeutung dies für die Interpretation der Ergebnisse hat, wird in Abschnitt 4.4 erläutert.

4.3 Berechnung der Gesamtwerte und Erstellung der Rankings

Der Abruf der Werte wurde am 6. Oktober 2010 durchgeführt. Durch den Abruf der Kennzahlen WIF, PageRank und Alexa Traffic Rank konnte jedem Websuchdienst der

¹³⁰ Für Mozilla Firefox unter <http://www.google.com/intl/de/toolbar/ff/index.html>, Abruf: 2010-09-12

¹³¹ <http://www.alexa.com/>, Abruf: 2010-09-12

¹³² Die Second Level Domain liegt im hierarchischem Domain Name System direkt unterhalb der Top Level Domain. Die Top Level Domain steht in einer URL rechts und die Second Level Domain links neben dem letzten Punkt. In dem Beispiel „<http://videos.google.com>“ stellt „google“ die Second Level Domain und „com“ die Top Level-Domain dar. Vgl dazu z. B. <http://www.denic.de/faq-single/335/245.html?cHash=05e6ed506e>, Abruf: 2010-10-03.

Grundgesamtheit ein Wert je Kennzahl zugeordnet werden.¹³³ Um mit Hilfe der Werte die Rankings der alternativen Websuchdienste zu erstellen, wird für jeden Websuchdienst ein Gesamtwert berechnet. Für diese Berechnung werden zuerst die Werte der drei Kennzahlen normiert, da die Werte für jede Kennzahl verschieden skaliert sind.

Dazu werden für jede Kategorie drei Rangfolgen¹³⁴ aufgestellt. In der ersten Rangfolge werden die Websuchdienste absteigend nach dem WIF sortiert, in der zweiten absteigend nach dem PageRank und in der dritten aufsteigend nach dem Alexa Traffic Rank. Für jede Rangfolge wird jedem Websuchdienst ein Rang zugeordnet. Der Rang entspricht der Position, die ein Websuchdienst in der jeweiligen Rangfolge einnimmt. Treten bei den Websuchdiensten gleiche Werte für die Kennzahlen auf, wird diesen Websuchdiensten als Rang das arithmetische Mittel der betroffenen Ränge zugeordnet. Auf diese Weise wurden für jeden Websuchdienst drei Ränge notiert. Diese Ränge stellen die normierten Werte für die drei Kennzahlen dar.

Durch die Bildung von Rängen wurden die normierten Werte der Kennzahlen in eine Ordinalskala gebracht. Diese Normierung wurde gewählt, da verschiedene Störfaktoren einen Einfluss auf die Werte der Kennzahlen haben können (vgl. Abschnitt 4.4). Durch die Normierung auf ordinal skalierte Werte, wirken sich die absoluten Abstände zwischen den Werten (und damit auch die Einflüsse durch Störfaktoren) weniger stark auf die Gesamtwerte aus.

Die Gesamtwerte wurden ermittelt, indem für jeden Websuchdienst aus den drei normierten Werten das arithmetische Mittel berechnet wurde.

Schließlich wurden die Rankings der Websuchdienste absteigend nach der Höhe des Gesamtwertes aufgestellt. Verfügen Websuchdienste einer Kategorie über den gleichen Gesamtwert, belegt der Websuchdienst mit dem höheren WIF den höheren Rang in der jeweiligen Kategorie. Die Rankings werden in Tab. 1 - 10 dargestellt und enthalten die ersten zehn Websuchdienste jeder Kategorie.¹³⁵

¹³³ Die Ergebnisse der folgenden Ausführungen finden sich in Tab. 1 – 10 wieder.

¹³⁴ An dieser Stelle wird anstatt des Begriffs Ranking der Begriff Rangfolge verwendet, um zwischen den Rangfolgen zur Normierung der Werte und den Rankings in Tab. 1 - 10 zu unterscheiden.

¹³⁵ Die Kategorie Personensuchdienste enthält nur sieben Suchdienste, da nur sieben Personensuchdienste für die Grundgesamtheit ermittelt werden konnten.

| | Name | URL | WIF | PageRank | Alexa Traffic Rank | Rang nach WIF | Rang nach PageRank | Rang nach Alexa Traffic Rank | Gesamtwert |
|-----|---------------------|---|----------|----------|--------------------|---------------|--------------------|------------------------------|------------|
| 1. | Flickr | http://www.flickr.com/ | 23839721 | 9 | 36 | 1 | 1 | 5 | 2,33 |
| 2. | Photobucket | http://photobucket.com/ | 9279912 | 7 | 84 | 2 | 5,5 | 7 | 4,83 |
| 3. | deviantART | http://www.deviantart.com/ | 856388 | 7 | 117 | 4 | 5,5 | 8 | 5,83 |
| 4. | SmugMug | http://www.smugmug.com/ | 1311179 | 7 | 1934 | 3 | 5,5 | 12 | 6,83 |
| 5. | TinyPic | http://www.tinypic.com/ | 240084 | 7 | 209 | 6 | 5,5 | 10 | 7,17 |
| 6. | Yahoo! Image Search | http://images.search.yahoo.com/ | 37621 | 6 | 4 | 12 | 13 | 2 | 9 |
| 7. | Bing Images | http://www.bing.com/images | 34325 | 6 | 22 | 13 | 13 | 3 | 9,67 |
| 8. | morgueFile | http://www.morguefile.com/ | 650970 | 7 | 9580 | 5 | 5,5 | 19 | 9,83 |
| 9. | TimEye | http://www.tineye.com/ | 44623 | 6 | 3302 | 11 | 13 | 14 | 12,67 |
| 10. | Picsearch | http://www.picsearch.com/ | 32139 | 7 | 11273 | 14 | 5,5 | 20 | 13,17 |

Tab. 1: Ranking der Bildersuchdienste

| | Name | URL | WIF | PageRank | Alexa Traffic Rank | Rang nach WIF | Rang nach PageRank | Rang nach Alexa Traffic Rank | Gesamtwert |
|-----|---------------------|---|----------|----------|--------------------|---------------|--------------------|------------------------------|------------|
| 1. | YouTube | http://www.youtube.com/ | 29575135 | 9 | 3 | 1 | 2 | 2 | 1,67 |
| 2. | MySpace Video | http://vids.myspace.com/ | 2453760 | 9 | 30 | 9 | 2 | 5 | 5,33 |
| 3. | Google Videos | http://video.google.com/ | 1284508 | 8 | 1 | 11 | 4 | 1 | 5,33 |
| 4. | Dailymotion | http://www.dailymotion.com/ | 12231721 | 7 | 106 | 3 | 8 | 9 | 6,67 |
| 5. | Aol Video | http://video.aol.com/ | 4073991 | 7 | 47 | 6 | 8 | 6 | 6,67 |
| 6. | Vimeo | http://www.vimeo.com/ | 672840 | 9 | 177 | 13 | 2 | 12 | 9 |
| 7. | Veoh | http://www.veoh.com/ | 5293294 | 7 | 456 | 5 | 8 | 15 | 9,33 |
| 8. | Truveo Video Search | http://www.truveo.com/ | 13260906 | 7 | 3656 | 2 | 8 | 20 | 10 |
| 9. | Metacafe | http://www.metacafe.com/ | 1695533 | 7 | 216 | 10 | 8 | 13 | 10,33 |
| 10. | Megavideo | http://www.megavideo.com/ | 1060852 | 6 | 83 | 12 | 16 | 7 | 11,67 |

Tab. 2: Ranking der Videosuchdienste

| | Name | URL | WIF | PageRank | Alexa Traffic Rank | Rang nach WIF | Rang nach PageRank | Rang nach Alexa Traffic Rank | Gesamtwert |
|-----|----------------------|---|--------|----------|--------------------|---------------|--------------------|------------------------------|------------|
| 1. | Mp3Raid | http://www.mp3raid.com/ | 75643 | 5 | 878 | 3 | 7 | 2 | 4 |
| 2. | BeeMP3 | http://beemp3.com/ | 59928 | 5 | 482 | 4 | 7 | 1 | 4 |
| 3. | FindSounds | http://www.findsounds.com/ | 47125 | 6 | 61795 | 5 | 2 | 7 | 4,67 |
| 4. | Mp3Bear | http://mp3bear.com/ | 22049 | 5 | 3917 | 6 | 7 | 3 | 5,33 |
| 5. | MP3 hunting | http://www.mp3hunting.com/ | 7372 | 5 | 9131 | 12 | 7 | 4 | 7,67 |
| 6. | seekasong | http://www.seekasong.com/ | 174272 | 4 | 65159 | 1 | 14,5 | 8 | 7,83 |
| 7. | MusicRobot | http://www.musicrobot.com/ | 7846 | 5 | 299222 | 11 | 7 | 10 | 9,33 |
| 8. | public domain sounds | http://www.pdsounds.org/ | 8864 | 5 | 551114 | 9 | 7 | 13 | 9,67 |
| 9. | Soungle | http://www.soungle.com/ | 2729 | 6 | 557012 | 14 | 2 | 14 | 10 |
| 10. | podscope | http://www.podscope.com/ | 8377 | 6 | 1201159 | 10 | 2 | 23 | 11,67 |

Tab. 3: Ranking der Audiosuchdienste

| | Name | URL | WIF | PageRank | Alexa Traffic Rank | Rang nach WIF | Rang nach PageRank | Rang nach Alexa Traffic Rank | Gesamtwert |
|----|-----------|---|---------|----------|--------------------|---------------|--------------------|------------------------------|------------|
| 1. | Intelius | http://search.intelius.com/ | 302565 | 6 | 1266 | 2 | 2,5 | 1 | 1,83 |
| 2. | 123people | http://www.123people.com/ | 4700007 | 6 | 1690 | 1 | 2,5 | 3 | 2,17 |
| 3. | Pipl | http://www.pipl.com/ | 29422 | 6 | 1521 | 4 | 2,5 | 2 | 2,83 |
| 4. | Wink | http://wink.com/ | 33998 | 6 | 17262 | 3 | 2,5 | 6 | 3,83 |
| 5. | PeekYou | http://www.peekyou.com/ | 25446 | 5 | 10199 | 5 | 6 | 5 | 5,33 |
| 6. | yasni | http://www.yasni.com/ | 1779 | 5 | 5190 | 6 | 6 | 4 | 5,33 |
| 7. | yoName | http://www.yoname.com/ | 870 | 5 | 292412 | 7 | 6 | 7 | 6,67 |

Tab. 4: Ranking der Personensuchdienste

| | Name | URL | WIF | PageRank | Alexa Traffic Rank | Rang nach WIF | Rang nach PageRank | Rang nach Alexa Traffic Rank | Gesamtwert |
|-----|------------------|---|----------|----------|--------------------|---------------|--------------------|------------------------------|------------|
| 1. | Yahoo! Answers | http://answers.yahoo.com/ | 1472296 | 8 | 4 | 2 | 1,5 | 1 | 1,5 |
| 2. | Answers.com | http://www.answers.com/ | 278705 | 8 | 142 | 3 | 1,5 | 2 | 2,17 |
| 3. | Answerbag | http://www.answerbag.com/ | 22532832 | 7 | 1407 | 1 | 3 | 4 | 2,67 |
| 4. | ChaCha | http://www.chacha.com/ | 39289 | 6 | 1868 | 4 | 5,5 | 7 | 5,5 |
| 5. | AllExperts | http://www.allexperts.com/ | 22830 | 6 | 1711 | 5 | 5,5 | 6 | 5,5 |
| 6. | Wikianswers | http://answers.wikia.com/wiki/Wikianswers | 1909 | 6 | 203 | 10 | 5,5 | 3 | 6,17 |
| 7. | Mahalo Answer | http://www.mahalo.com/answers | 12259 | 5 | 1545 | 6 | 9,5 | 5 | 6,83 |
| 8. | Ask Me Help Desk | http://www.askmehelpdesk.com/ | 2258 | 6 | 6872 | 9 | 5,5 | 8 | 7,5 |
| 9. | FunAdvice | http://www.funadvice.com/ | 8011 | 5 | 11219 | 7 | 9,5 | 9 | 8,5 |
| 10. | QuestionBin | http://www.questionbin.com/ | 6323 | 5 | 89185 | 8 | 9,5 | 13 | 10,17 |

Tab. 5: Ranking der Frage-Antwort-Dienste

| | Name | URL | WIF | PageRank | Alexa Traffic Rank | Rang nach WIF | Rang nach PageRank | Rang nach Alexa Traffic Rank | Gesamtwert |
|-----|------------------------|---|------------|-----------------|---------------------------|----------------------|---------------------------|-------------------------------------|-------------------|
| 1. | Digg | http://digg.com/ | 30364573 | 8 | 115 | 2 | 4 | 1 | 2,33 |
| 2. | Mister Wong | http://www.mister-wong.de/ | 39270657 | 9 | 2085 | 1 | 1 | 7 | 3 |
| 3. | StumbleUpon | http://www.stumbleupon.com/ | 21681633 | 8 | 170 | 3 | 4 | 2 | 3 |
| 4. | Mixx | http://www.mixx.com/ | 8060092 | 8 | 1197 | 4 | 4 | 5 | 4,33 |
| 5. | reddit | http://www.reddit.com/ | 2545418 | 8 | 245 | 6 | 4 | 3 | 4,33 |
| 6. | Delicious | http://www.delicious.com/ | 1444187 | 7 | 268 | 7 | 7 | 4 | 6 |
| 7. | Diigo | http://www.diigo.com/ | 5639366 | 6 | 1556 | 5 | 8,5 | 6 | 6,5 |
| 8. | Connotea | http://www.connotea.org/ | 127962 | 8 | 4574 | 8 | 4 | 9 | 7 |
| 9. | Clipmarks | http://www.clipmarks.com/ | 8661 | 6 | 3529 | 11 | 8,5 | 8 | 9,17 |
| 10. | Social-Bookmarking_Net | http://www.social-bookmarking.net/ | 29975 | 4 | 11171 | 10 | 12,5 | 10 | 10,83 |

Tab. 6: Ranking der Social Bookmarking-Dienste

| | Name | URL | WIF | PageRank | Alexa Traffic Rank | Rang nach WIF | Rang nach PageRank | Rang nach Alexa Traffic Rank | Gesamtwert |
|-----|--------------------|---|----------|----------|--------------------|---------------|--------------------|------------------------------|------------|
| 1. | Technorati | http://www.technorati.com/ | 21577329 | 8 | 897 | 1 | 1,5 | 3 | 1,83 |
| 2. | Topix | http://www.topix.com/ | 1186208 | 8 | 838 | 6 | 1,5 | 2 | 3,17 |
| 3. | Google Blog Search | http://blogsearch.google.com/ | 661457 | 7 | 1 | 7 | 5 | 1 | 4,33 |
| 4. | Blogarama | http://www.blogarama.com/ | 6692157 | 7 | 22492 | 2 | 5 | 8 | 5 |
| 5. | Bloggernity | http://www.bloggernity.com/ | 2580554 | 7 | 28764 | 4 | 5 | 9 | 6 |
| 6. | Blog Catalog | http://www.blogcatalog.com/ | 3048695 | 4 | 910 | 3 | 13,5 | 4 | 6,83 |
| 7. | IceRocket | http://www.icerocket.com/ | 146374 | 7 | 5590 | 12 | 5 | 6 | 7,67 |
| 8. | Best of the Web | http://blogs.botw.org/ | 123708 | 7 | 4599 | 13 | 5 | 5 | 7,67 |
| 9. | Bloghub | http://www.bloghub.com/ | 2173623 | 6 | 41114 | 5 | 9 | 11 | 8,33 |
| 10. | Blog Search Engine | http://www.blogsearchengine.com/ | 561120 | 6 | 40390 | 9 | 9 | 10 | 9,33 |

Tab. 7: Ranking der Blogsuchdienste

| | Name | URL | WIF | PageRank | Alexa Traffic Rank | Rang nach WIF | Rang nach PageRank | Rang nach Alexa Traffic Rank | Gesamtwert |
|-----|--------------------------|---|------------|-----------------|---------------------------|----------------------|---------------------------|-------------------------------------|-------------------|
| 1. | Twitter Search | http://search.twitter.com/ | 386617 | 8 | 9 | 2 | 1 | 2 | 1,67 |
| 2. | TweetMeme | http://tweetmeme.com/ | 444353 | 7 | 381 | 1 | 2,5 | 3 | 2,17 |
| 3. | Twingly Microblog Search | http://www.twingly.com/microblogsearch | 70000 | 7 | 19523 | 4 | 2,5 | 6 | 4,17 |
| 4. | Twellow | http://www.twellow.com/ | 91133 | 6 | 8668 | 3 | 6,5 | 5 | 4,83 |
| 5. | monitter | http://www.monitter.com/ | 31428 | 6 | 75161 | 5 | 6,5 | 10 | 7,17 |
| 6. | Twitscoop | http://www.twitscoop.com/ | 15841 | 6 | 48715 | 7 | 6,5 | 9 | 7,5 |
| 7. | BackTweets | http://backtweets.com/ | 19323 | 5 | 21605 | 6 | 12,5 | 7 | 8,5 |
| 8. | Tweet Scan | http://www.tweetscan.com/ | 3282 | 6 | 121577 | 9 | 6,5 | 15 | 10,17 |
| 9. | Twitterfall | http://www.twitterfall.com/ | 1326 | 6 | 80088 | 14 | 6,5 | 11 | 10,5 |
| 10. | TweepSearch | http://tweepsearch.com/ | 3589 | 5 | 82826 | 8 | 12,5 | 12 | 10,83 |

Tab. 8: Ranking der Twittersuchdienste

| | Name | URL | WIF | PageRank | Alexa Traffic Rank | Rang nach WIF | Rang nach PageRank | Rang nach Alexa Traffic Rank | Gesamtwert |
|-----|--------------------|---|----------|----------|--------------------|---------------|--------------------|------------------------------|------------|
| 1. | Google News | http://news.google.com/ | 3909933 | 8 | 1 | 2 | 2,5 | 1 | 1,83 |
| 2. | Topix | http://www.topix.com/ | 1184579 | 8 | 838 | 4 | 2,5 | 7 | 4,5 |
| 3. | Wikio | http://www.wikio.com/ | 1367969 | 8 | 2799 | 3 | 2,5 | 12 | 5,83 |
| 4. | AOL News | http://www.aolnews.com/ | 15864876 | 6 | 661 | 1 | 12 | 5 | 6 |
| 5. | Lycos News | http://news.lycos.com/ | 440714 | 7 | 1628 | 6 | 7 | 8 | 7 |
| 6. | NewsNow | http://www.newsnow.co.uk/ | 1046204 | 7 | 1995 | 5 | 7 | 11 | 7,67 |
| 7. | World News | http://wn.com/ | 161007 | 7 | 1718 | 8 | 7 | 9 | 8 |
| 8. | Bing News | http://www.bing.com/news | 312326 | 5 | 22 | 7 | 18,5 | 3 | 9,5 |
| 9. | Yahoo! News Search | http://news.search.yahoo.com/ | 15562 | 5 | 4 | 14 | 18,5 | 2 | 11,5 |
| 10. | PressDisplay | http://www.pressdisplay.com/ | 40628 | 7 | 33528 | 12 | 7 | 17 | 12 |

Tab. 9: Ranking der Nachrichtensuchdienste

| | Name | URL | WIF | PageRank | Alexa Traffic Rank | Rang nach WIF | Rang nach PageRank | Rang nach Alexa Traffic Rank | Gesamtwert |
|-----|---------------------------|---|--------|----------|--------------------|---------------|--------------------|------------------------------|------------|
| 1. | Google Scholar | http://scholar.google.com/ | 488890 | 9 | 1 | 1 | 2 | 1 | 1,33 |
| 2. | Scirus | http://www.scirus.com/ | 181591 | 9 | 37579 | 4 | 2 | 6 | 4 |
| 3. | CiteULike | http://www.citeulike.org/ | 269405 | 8 | 13077 | 2 | 6,5 | 5 | 4,5 |
| 4. | Science.gov | http://www.science.gov/ | 248416 | 9 | 709615 | 3 | 2 | 11 | 5,33 |
| 5. | INFOMINE | http://infomine.ucr.edu/ | 18390 | 8 | 11966 | 8 | 6,5 | 4 | 6,17 |
| 6. | CiteSeerX | http://citeseerx.ist.psu.edu/ | 8511 | 8 | 2509 | 9 | 6,5 | 3 | 6,17 |
| 7. | WorldWideScience | http://worldwidescience.org/ | 130264 | 8 | 424202 | 5 | 6,5 | 9 | 6,83 |
| 8. | vascoda | http://www.vascoda.de/ | 81750 | 8 | 1055117 | 7 | 6,5 | 15 | 9,5 |
| 9. | Microsoft Academic Search | http://academic.research.microsoft.com/ | 6475 | 6 | 23 | 11 | 17 | 2 | 10 |
| 10. | BASE | http://www.base-search.net/ | 7908 | 7 | 581195 | 10 | 12 | 10 | 10,67 |

Tab. 10: Ranking der Wissenschaftssuchdienste

4.4 Interpretation der Rankings

In diesem Abschnitt wird erläutert, welche Informationen den aufgestellten Rankings in Tab. 1 - 10 entnommen werden können und welche Aspekte bei der Interpretation der enthaltenen Ergebnisse beachtet werden müssen.

Aus den Rankings kann der ermittelte Rang nach Gesamtwert, der Name, die untersuchte URL, die abgerufenen Werte der Kennzahlen und der Rang nach WIF, PageRank und Alexa Traffic Rank sowie der berechnete Gesamtwert der ersten zehn Websuchdienste der jeweiligen Kategorie abgelesen werden.

Eine Interpretation der Gesamtwerte ist dadurch begrenzt, dass sich die Gesamtwerte gleich gewichtet aus den ordinal skalierten Werten der Kennzahlen zusammensetzen. Die Abstände und Quotienten von ordinal skalierten Werten, können nicht interpretiert werden [BaVo08, 9]. Deshalb dienen auch die Gesamtwerte nur dazu, die Reihenfolge der Websuchdienste in den Kategorien zu bestimmen.

Bei einer Interpretation der abgerufenen Werte der Kennzahlen sind folgende Aspekte zu beachten.

Wie in Abschnitt 4.2 erörtert wurde, kann es vorkommen, dass ein Websuchdienst über verschiedene URLs erreichbar ist. Daher können auch die Verlinkungen auf den Websuchdienst durch verschiedene URLs angegeben worden sein. In dieser Arbeit wurde nur eine URL je Websuchdienst verwendet, um die Werte der Kennzahlen abzurufen. Deshalb sind die Werte der Kennzahlen und die Rankings davon abhängig, welche URL für deren Abruf verwendet wurde.

Damit könnte z. B. erklärt werden, warum die Bildersuche von Google¹³⁶ in diesem Ranking nicht unter den zehn populärsten Bildersuchdiensten erscheint.¹³⁷ Die Google Bildersuche wurde in dieser Arbeit mit der URL <http://www.google.com/imghp> ausgewertet. Diese erreichte am 6. Oktober 2010 für den WIF einen Wert von 7881. Alternativ ist die Google Bildersuche auch unter der URL <http://images.google.com/> erreichbar. Für

¹³⁶ Die Bildersuche von Google ist der meistgenutzte alternative Websuchdienst im WWW. Vgl. <http://www.at-web.de/blog/20101016/suchmaschinen-ranking-in-deutschland-september-2010.htm>, Abruf: 2010-10-13

¹³⁷ In dem folgenden Abschnitt werden zwei Hypothesen aufgestellt, die sich im Rahmen dieser Arbeit nicht überprüfen lassen, da eine Vielzahl von Faktoren einen Einfluss auf die Anzahl von externen Inlinks haben können.

diese URL betrug der Wert des WIF zum gleichen Zeitpunkt 289840. Daher hätte die Google Bildersuche in diesem Ranking einen höheren Wert für den WIF erreicht, wenn die URL <http://images.google.com/> für den Abruf der Kennzahl verwendet worden wäre.

Ebenso wäre es möglich, dass ein Websuchdienst, der unter mehreren URLs verfügbar ist, einen geringeren Wert für den WIF erhält als ein Websuchdienst, der nur unter einer URL verfügbar ist, wenn die Gesamtanzahl von Verlinkungen auf die unterschiedlichen URLs verteilt ist. Da nur eine URL für den Abruf des WIF verwendet wurde, kann es also sein, dass nur ein Teil der Verlinkungen durch den Wert des WIF abgebildet werden.

Diese Probleme lassen sich allerdings nicht in der Berechnung der Kennzahl berücksichtigen, da zum einen nicht bestimmt werden kann, unter wie vielen URLs ein Websuchdienst erreichbar ist und zum anderen nicht, aus welchen Gründen mehrere URLs angeboten werden.¹³⁸

Eine weitere Unsicherheit bezüglich der Werte ergibt sich daraus, dass unter der URL, über die ein Websuchdienst erreichbar ist, weitere Dienste angeboten werden können. Deshalb muss eine Verlinkung auf eine untersuchte URL nicht zwangsläufig einem Websuchdienst gelten.

Beispiele dafür sind die Communities und Plattformen. Diese Communities und Plattformen belegen in dem Ranking der Bildersuchdienste und in dem Ranking der Videosuchdienste vordere Ränge. Neben der Suche werden die Communities und Plattformen auch zur Speicherung und Verwaltung von Dokumenten sowie zur Kommunikation mit anderen Nutzern verwendet.

Dadurch kann nicht bestimmt werden, welcher Dienst der URL, welchen Einfluss auf die Werte der Kennzahlen WIF und PageRank hat.

Bei dem Alexa Traffic Rank lässt sich dieses Problem ebenfalls feststellen. Der Alexa Traffic Rank eines Websuchdienstes bezieht sich immer auf die Second Level Domain der URL des Websuchdienstes, weil für den Alexa Traffic Rank alle Nutzungsdaten ausgewertet werden, die unter der Second Level Domain einer URL gemessen werden. Dadurch kann nicht festgestellt werden, welcher Dienst der Second Level Domain, welchen Einfluss auf den Alexa Traffic Rank hat.

¹³⁸ Ein Grund könnte z. B. ein Wechsel der URL sein. Für den Zeitraum der Umstellung könnten mehrere URLs angeboten werden.

In den Rankings kann sich dieses Problem vor allem positiv für die Websuchdienste auswirken, bei denen sehr populäre Dienste unter der Second Level Domain mit angeboten werden. So kann es z. B. sein, dass die alternativen Suchdienste von Google, Bing und Yahoo von einem hohen Alexa Traffic Rank profitieren, der sich überwiegend aus den Nutzungsdaten bezüglich der allgemeinen Suchmaschine der Anbieter ergibt.¹³⁹

Die Anzahl der externen Inlinks (WIF) wurde durch den Crawler des Yahoo Site Explorer ermittelt.¹⁴⁰ Auf Grund der Struktur des WWW (vgl. Abschnitt 2.2.1) kann ein Crawler nicht alle im WWW gesetzten Verlinkungen auf die Webseite eines Websuchdiensten erfassen. Deshalb stellt die in dieser Arbeit ermittelte Anzahl von externen Inlinks nicht die tatsächliche Anzahl von Verlinkungen auf einen Websuchdienst dar.

Wie in Abschnitt 2.4.1 erörtert wurde, können die Gründe für die Erstellung einer Verlinkung andere sein, als eine Empfehlung der Webseite. Vor allem die subjektiven Maßnahmen zur Erhöhung der Sichtbarkeit der eigenen Website (z. B. durch SEO) werden nicht zwangsläufig durch die automatischen Verfahren, die die Daten für die Werte der Kennzahlen sammeln, erkannt. Deshalb können die Werte der Kennzahlen von solchen Maßnahmen beeinflusst sein. Welche der Websuchdienste in den Rankings beeinflusste Werte aufweisen, kann allerdings nicht nachgewiesen werden.

Um feststellen zu können, auf welchen Zeitpunkt oder Zeitraum sich die abgerufenen Werte der Kennzahlen beziehen, muss bestimmt werden, zu welchem Zeitpunkt oder für welchen Zeitraum die Werte der Kennzahlen berechnet worden sind.

Die Werte des WIF bilden den Zeitpunkt ab, an dem der Crawler des Yahoo Site Explorers zuletzt seinen Index aktualisiert hat. Da dies ein sich ständig wiederholender Prozess ist, stellt der Abruf des WIF eine Momentaufnahme dar.

Die Daten für den PageRank, der in der Google Toolbar angezeigt wird, werden von Google nur alle drei bis vier Monate aktualisiert, da dies ein aufwändiger Prozess ist.¹⁴¹ Die in dieser Arbeit abgerufenen Werte für den PageRank wurden also maximal vor vier Monaten berechnet.

¹³⁹ Auch diese Hypothese kann nicht weiter überprüft werden.

¹⁴⁰ Vgl. <http://help.yahoo.com/l/us/yahoo/search/siteexplorer/siteexplorer-01.html>, Abruf: 2010-10-13

¹⁴¹ Vgl. <http://www.mattcutts.com/blog/more-info-on-pagerank/>, Abruf: 2010-10-13

Der Alexa Traffic Rank wird für jeden Tag aus den aufgezeichneten Nutzungsdaten der letzten drei Monate berechnet.¹⁴² Zum Zeitpunkt des Abrufs ist der Alexa Traffic Rank also aktuell, bezieht sich allerdings auf einen Zeitraum von drei Monaten.

Die Gesamtwerte der Websuchdienste enthalten also einen Wert (den des PageRanks), der maximal vor vier Monaten berechnet wurde.

Schließlich ergibt sich aus der verwendeten Methode zur Ermittlung der Grundgesamtheit, dass in den Rankings nur englisch- oder deutschsprachige Websuchdienste enthalten sind. Im WWW können allerdings weitere alternative Websuchdienste in anderen Sprachen existieren, die trotz einer hohen Popularität nicht in diesen Rankings erscheinen.

¹⁴² Vgl. http://www.alexa.com/help/traffic_learn_more, Abruf: 2010-10-13

5 Zusammenfassung und Ausblick

Ziel dieser Arbeit war es, einen Überblick über alternative Websuchdienste zum derzeitigen Stand zu geben. Dazu wurde gezeigt, welche unterschiedlichen Suchansätze Websuchdienste verfolgen können. Daraus konnte wiederum abgeleitet werden, dass eine allgemeingültige Kategorisierung für alternative Websuchdienste nicht möglich ist, weil in den Konzepten der alternativen Websuchdienste die Suchansätze verschieden kombiniert und variiert werden können. Für den Vergleich alternativer Websuchdienste wurden zehn Kategorien von alternativen Websuchdiensten ausgewählt und näher erläutert. Mit den webometrischen Kennzahlen, die für den Vergleich ausgewählt wurden, sind eine Reihe von Nachteilen bezüglich der Validität und Reliabilität verbunden. Trotzdem konnten diese verwendet werden, um die populärsten alternativen Websuchdienste zu ermitteln. Die zehn populärsten Websuchdienste jeder ausgewählten Kategorie wurden in Rankings dargestellt. Die Rankings bilden eine Momentaufnahme der derzeit populärsten alternativen Websuchdienste ab.

Diese Momentaufnahme kann zukünftigen Untersuchungen zu alternativen Websuchdiensten oder Untersuchungen zu dem Markt der Websuchdienste dienen. Mit einer gleichen Arbeit zu einem späteren Zeitpunkt könnte z. B. überprüft werden, inwieweit sich das Angebot an alternativen Websuchdiensten verändert hat. Damit könnte festgestellt werden, welche alternativen Websuchdienste aus den Rankings ausgeschieden sind, welche weiterhin darin enthalten sind und wie sich die Positionen von Websuchdiensten in den Rankings verändert haben.

Im Gegensatz zu dem Vorgehen in dieser Arbeit könnten die Rankings in einer weiteren Arbeit auch durch eine Umfrage über die Popularität der alternativen Websuchdienste aufgestellt werden. Dadurch könnten sich die Ergebnisse der Arbeiten gegenseitig ergänzen.

Als eine Grenze ist in dieser Arbeit die Tatsache zu nennen, dass keine vollständige Abbildung aller alternativen Websuchdienste geleistet werden konnte. Dieses ist zum einen auf die verwendete Methodik zurückzuführen und zum anderen darauf, dass die Anzahl von verfügbaren Websuchdiensten nicht bestimmt werden kann.

Eine Schwachstelle der Rankings ist die Aussagekraft der Werte. Dieses ist zum einen auf die Kennzahlen selbst und zum anderen auf die Berechnung der Gesamtwerte zurückzuführen. Deshalb müssen Aussagen, die aus einem der Rankings abgeleitet werden, immer im Zusammenhang mit diesem Ranking erläutert werden.

In Bezug auf zukünftige Entwicklungen bleibt es weiterhin spannend, inwieweit alternative Websuchdienste bezüglich der Qualität der Suchergebnisse an die allgemeinen Suchmaschinen anschließen können und welche Position diese gegenüber den allgemeinen Suchmaschinen auf dem Markt für Websuchdienste einnehmen werden. Dabei ist abzuwarten, wie sich bestehende Suchansätze weiterentwickeln und welche neuen Suchansätze entstehen werden. Besonders aufmerksam sind die Entwicklungen der Echtzeitsuche, der Personalisierten Suche, der Lokalen Suche und der Semantischen Suche zu beobachten, da diese vielversprechende Ansätze verfolgen. Neue Suchansätze können z.B. auch auf Grund von neuen Entwicklungen im WWW oder durch ein verändertes Suchverhalten der Nutzer entstehen.

Für Methoden zur Bewertung der Popularität von alternativen Websuchdiensten ist es wünschenswert, dass Kennzahlen entwickelt werden, die weniger beeinflussbar sind und ein höheres Maß an Validität und Reliabilität besitzen, als die in dieser Arbeit verwendeten.

Literaturverzeichnis

- [Alln97] Almind, T.; Ingwersen, P.: Informetric analyses on the World Wide Web: A methodological approach to "webometrics". In: Journal of Documentation. Nr. 53 (4), 1997, S. 404-426.
- [BaRi99] Baeza-Yates, R.; Ribeiro-Neto, B.: Modern Information Retrieval. ACM Press, New York, 1999.
- [BaVo08] Bankhofer, U.; Vogel, J.: Datenanalyse und Statistik: Eine Einführung für Ökonomen im Bachelor. Gabler Verlag, Wiesbaden, 2008.
- [Beka09] Bekavac, B.: Schweizer Portal für Recherche im Internet (SPRINT). 2009, <http://sprint.informationswissenschaft.ch/suchdienste/>, Abruf: 2010-07-05.
- [Bern89] Berners-Lee, T.: Information Management: A Proposal. 1989, <http://www.w3.org/History/1989/proposal.html>, Abruf: 2010-07-16.
- [BHLa01] Berners-Lee, T.; Hendler, J.; Lassila, O.: The Semantic Web. 2001, <http://www.scientificamerican.com/article.cfm?id=the-semantic-web&page=2>, Abruf: 2010-09-13.
- [BjIn04] Björneborn, L.; Ingwersen, P.: Toward a basic framework for webometrics. In: Journal of the American Society for Information Science and Technology. Nr. 55 (14), 2004, S. 1216-1227.
- [Björ04] Björneborn, L.: Small-World Link Structures across an Academic Web Space: A Library and Information Science Approach . Dissertation vorgelegt an der Department of Information Studies, Royal School of Library and Information Science, Denmark. Kopenhagen, 2004.
- [BKMR+00] Broder, A.; Kumar, R.; Maghoul, F.; Raghavan, P.; Rajagopalan, S.; Stata, R.; Tomkins, A.; Wiener, J.: Graph structure in the Web. 2000, <http://www9.org/w9cdrom/160/160.html>, Abruf: 2010-07-16.
- [Brod02] Broder, A.: A taxonomy of web search. 2002, www.sigir.org/forum/F2002/broder.pdf, Abruf: 2010-07-19.

- [BrPa98] Brin, S.; Page, L.: The Anatomy of a Large-Scale Hypertextual Web Search Engine. 1998, <http://ilpubs.stanford.edu:8090/361/1/1998-8.pdf>, Abruf: 2010-08-03.
- [CrHa09] Craswell, N.; Hawking, D.: Web Information Retrieval. In: Ayse Göker, John Davies (Hrsg.): Information Retrieval: Searching the 21st Century. John Wiley & Sons, Ltd., West Sussex et al., 2009, S. 85-99.
- [Dopi09] Dopichaj, P.: Ranking-Verfahren für Web-Suchmaschinen. In: Dirk Lewandowski (Hrsg.): Handbuch Internet-Suchmaschinen. Aka Verlag, Heidelberg, 2009, S. 101-115.
- [Ferb03] Ferber, R.: Information Retrieval: Suchmodelle und Data-Mining-Verfahren für Textsammlungen und das Web. dpunkt.verlag, Heidelberg, 2003.
- [Garf79] Garfield, E.: Citation Indexing - Its Theory and Application in Science, Technology, and Humanities. John Wiley & Sons Inc, o. O., 1979.
- [GBRi09] Griesbaum, J.; Bekavac, B.; Rittberger, M.: Typologie der Suchdienste im Internet. In: Dirk Lewandowski (Hrsg.): Handbuch Internet-Suchmaschinen. Aka Verlag, Heidelberg, 2009, S. 18-52.
- [Gele03] Gelernter, J.: At the limits of Google: Specialized search engines. 2003, <http://www.allbusiness.com/technology/software-services-applications-search-engines/10603593-1.html>, Abruf: 2010-08-06.
- [GrBe04a] Griesbaum, J.; Bekavac, B.: Web-Suche im Umbruch? Entwicklungstendenzen bei Web-Suchdiensten. 2004, http://www.joachim-griesbaum.de/files/Entwicklungstendenzen_bei_Websuchdiensten_volltext.pdf, Abruf: 2010-07-22.
- [GrBe04b] Griesbaum, J.; Bekavac, B.: Von der Kommerzialisierung bis zum Deep Web: Problemfelder der Internetsuche. In: Rainer Hammwöhner, Marc Rittberger, Wolfgang Semar (Hrsg.): Wissen in Aktion. Der Primat der Pragmatik als Motto der Konstanzer Informationswissenschaft. UVK Verlagsgesellschaft mbH, Konstanz, 2004, S. 35-50.
- [GrFr04] Grossman, D.; Frieder, O.: Information Retrieval : Algorithms and Heuristics. Springer Netherland, Berlin, 2004.

- [Grie07] Griesbaum, J.: Entwicklungstrends im Web Information Retrieval: Neue Potentiale für die Webrecherche durch Personalisierung & Web 2.0-Technologien. 2007, http://www.web-information-retrieval.de/files/volltext_kollaboratives_retrievalfinal_040707.pdf, Abruf: 2010-07-22.
- [HaHe06] Hassan-Montero, Y.; Herrero-Solana, V.: Improving Tag-Clouds as Visual Information Retrieval Interfaces. 2006, http://www.nosolousabilidad.com/hassan/improving_tagclouds.pdf, Abruf: 2010-08-05.
- [Heck08] Heckner, M.: Tagging, Rating, Posting: Studying Forms of User Contribution for Web-based Information Management and Information Retrieval. Dissertation vorgelegt an der Philosophische Fakultät IV - Sprach und Literaturwissenschaften - der Universität Regensburg. Regensburg, 2008.
- [HMSi02] Henzinger, M.; Motwani, R.; Silverstein, C.: Challenges in Web Search Engines. 2002, <http://www.acm.org/sigs/sigir/forum/F2002/henzinger.pdf>, Abruf: 2010-07-07.
- [Höch09] Höchstötter, N.: Methoden der Erhebung von Nutzerdaten und ihre Anwendung in der Suchmaschinenforschung. In: Dirk Lewandowski (Hrsg.): Handbuch Internet-Suchmaschinen. Aka Verlag, Heidelberg, 2009, S. 175-203.
- [Ingw98] Ingwersen, P.: The Calculation of Web Impact Factors. In: Journal of Documentation. Nr. 54 (2), 1998, S. 236-243.
- [IVWR09] o. V.: Anlage 1 zu den IVW-Richtlinien für Online-Angebote. Definitionen und technische Erläuterungen. 2009, http://daten.ivw.eu/download/pdf/Anlage1_Richtlinien_v2_2.pdf, Abruf: 2010-10-01.
- [Karo07] Karow, M.: Personalisierte Suche. In: Heinz Lothar Grob, Gottfried Vossen (Hrsg.): Entwicklungen im Web 2.0 aus technischer, ökonomischer und sozialer Sicht. Arbeitsberichte des Kompetenzzentrums Internetökonomie und Hybridität Nr. 51, Münster, 2007, S. 153-167.

- [Karz10] Karzauninkat, S.: Suchfibel. o. J., <http://www.suchfibel.de/4spez/personen-suchen.html>, Abruf: 2010-08-10.
- [KNMe08] Kalashnikov, D; Nuray-Turan, R.; Mehrotra, S.: Towards Breaking the Quality Curse. A Web-Querying Approach to Web People Search. 2008, http://www.ics.uci.edu/~dvk/pub/SIGIR08_dvk.pdf, Abruf: 2010-08-10.
- [LaMe06] Langville, A.; Meyer, C.: Google's PageRank and Beyond: The Science of Search Engine Rankings. Princeton University Press, Princeton, 2006.
- [Lars96] Larson, R.: Bibliometrics of the World Wide Web: An Exploratory Analysis of the Intellectual Structure of Cyberspace. 1996, <https://sherlock.ischool.berkeley.edu/asis96/asis96.html>, Abruf: 2010-08-02.
- [LeHö07] Lewandowski, D.; Höchstötter, N.: Qualitätsmessung bei Suchmaschinen - System und nutzerbezogene Evaluationsmaße. 2007, http://www.bui.haw-hamburg.de/fileadmin/user_upload/lewandowski/doc/Qualitaetsmessung_bei_Suchmaschinen_Informatik_Spektrum_2007.pdf, Abruf: 2010-09-30.
- [LeHö08] Lewandowski, D.; Höchstötter, N.: Wie effektiv sind Suchmaschinen zur Recherche nach Bildern von berühmten Persönlichkeiten?. 2008, http://www.bui.haw-hamburg.de/fileadmin/user_upload/lewandowski/doc/Bildersuche_Preprint.pdf, Abruf: 2010-07-22.
- [LeMa06] Lewandowski, D.; Mayr, P.: Exploring the Academic Invisible Web. 2006, http://www.bui.haw-hamburg.de/fileadmin/user_upload/lewandowski/doc/Exploring_the_Academic_Invisible_Web.pdf, Abruf: 2010-09-14.
- [Lewa04a] Lewandowski, D.: Datumsbeschränkung bei WWW-Suchanfragen: Eine Untersuchung der Möglichkeiten der zeitlichen Einschränkung von Suchanfragen in den Suchmaschinen Google, Teoma und Yahoo. 2004, http://www.bui.haw-hamburg.de/fileadmin/user_upload/lewandowski/doc/Datumsbeschraenkung_Suchmaschinen.pdf, Abruf: 2010-09-15.
- [Lewa04b] Lewandowski, D.: Bewertung von linktopologischen Verfahren als bestimmender Ranking-Faktor bei WWW-Suchmaschinen. 2004, http://www.bui.haw-hamburg.de/fileadmin/user_upload/lewandowski/

- doc/Linktopologische_Verfahren_bestimmender_Rankingfaktor.pdf,
Abruf: 2010-08-02.
- [Lewa05] Lewandowski, D.: Web Information Retrieval: Technologien zur Informationssuche im Internet. Frankfurt am Main: DGI, Frankfurt am Main, 2005.
- [Lewa08] Lewandowski, D.: Suchmaschinenforschung im Kontext einer zukünftigen Webwissenschaft. 2008, http://web.bui.haw-hamburg.de/fileadmin/user_upload/lewandowski/doc/Suchmaschinenforschung_Kontext_Webwissenschaft.pdf, Abruf: 2010-09-28.
- [Lewa09a] Lewandowski, D.: Einstieg in Real Time Search. 2009, http://www.bui.haw-hamburg.de/fileadmin/user_upload/lewandowski/doc/Real_Time_Suche_Lewandowski.pdf, Abruf: 2010-07-22.
- [Lewa09b] Lewandowski, D.: Wie Suchmaschinen von Social Software profitieren. 2009, http://www.bui.haw-hamburg.de/fileadmin/user_upload/lewandowski/doc/SSAW_Lewandowski.pdf, Abruf: 2010-09-25.
- [Lewa09c] Lewandowski, D.: Spezialsuchmaschinen. In: Dirk Lewandowski (Hrsg.): Handbuch Internet-Suchmaschinen. Aka Verlag, Heidelberg, 2009, S. 53-69.
- [LKHo05] Lee, S.; Kim, S.; Hong, S.: On URL Normalization. 2005, <http://dblab.ssu.ac.kr/publication/LeKi05a.pdf>, Abruf: 2010-10-01.
- [LoSe06] Lo, B.; Sedhain, R.: How reliable are website rankings? Implications for e-business advertising and internet search. 2006, http://www.iacis.org/iis/2006_iis/PDFs/Lo_Sedhain.pdf, Abruf: 2010-09-30.
- [MaLe09] Maaß, C.; Lewandowski, D.: Frage-Antwort-Dienste als alternativer Suchansatz?. 2009, http://www.bui.haw-hamburg.de/fileadmin/user_upload/lewandowski/doc/Frage_Antwort_Dienste_als_alternativer_Suchansatz.pdf, Abruf: 2010-07-22.
- [MRSc08] Manning, C.; Raghavan, P.; Schütze, H.: Introduction to Information Retrieval. Cambridge University Press, Cambridge, 2008.

- [MSHP09] Maaß, C.; Skusa, A.; Heß, A.; Pietsch, G.: Der Markt für Internet-Suchmaschinen. In: Dirk Lewandowski (Hrsg.): Handbuch Internet-Suchmaschinen. Aka Verlag, Heidelberg, 2009, S. 3-17.
- [O'Rei05] O'Reilly, T.: What is Web 2.0?. 2005, <http://www.oreilly.de/artikel/web20.html>, Abruf: 2010-07-22.
- [PBMW98] Page, L.; Brin, S.; Motwani, R.; Winograd, T.: The PageRank Citation Ranking: Bringing Order to the Web. 1998, <http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf>, Abruf: 2010-09-10.
- [Prit69] Pritchard, A.: Statistical bibliography or bibliometrics?. In: Journal of Documentation. Nr. 25 (4), 1969, S. 348-349.
- [RiBr09] Riemer, K.; Brüggemann, F.: Personalisierung der Internetsuche - Lösungstechniken und Marktüberblick. In: Dirk Lewandowski (Hrsg.): Handbuch Internet-Suchmaschinen. Aka Verlag, Heidelberg, 2009, S. 148-171.
- [Rous97] Ronald Rousseau: Sitations: an exploratory study. 1997, <http://www.cindoc.csic.es/cybermetrics/articles/v1i1p1.pdf>, Abruf: 2010-08-02.
- [SaMc87] Salton, G.; McGill, M. (Übersetzung: von Keitz, W.): Information Retrieval - Grundlegendes für Informationswissenschaftler. McGraw-Hill, Hamburg et al., 1987.
- [ShPr01] Sherman, C.; Price, G.: The Invisible Web: Uncovering Information Sources Search Engines Can't See. Information Today, Medford, 2001.
- [SkMa08] Skusa, A.; Maaß, C.: Suchmaschinen: Status quo und Entwicklungstendenzen. In: Dirk Lewandowski, Christian Maaß (Hrsg.): Web-2.0-Dienste als Ergänzung zu algorithmischen Suchmaschinen. Logos, Berlin, 2008, S. 1-11.
- [Stoc07] Stock, W.: Information Retrieval: Informationen suchen und finden. Oldenbourg, München, Wien, 2007.
- [Thel04] Thelwall, M.: Link Analysis: An Information Science Approach. Academic Press, San Diego et al., 2004.
- [Thel09] Thelwall, M.: Introduction to Webometrics: Quantitative Web Research for the Social Sciences. Morgan & Claypool, San Rafael, 2009.

- [ThHa07] Thelwall, M.; Hasler, L.: Blog Search Engines. 2007, http://cybermetrics.wlv.ac.uk/paperdata/BlogSearchEngines_preprint.doc, Abruf: 2010-09-15.
- [ThRu06] Thelwall, M.; Ruschenburg, T.: Grundlagen und Forschungsfelder der Webometrie. 2006, http://www.phil-fak.uni-duesseldorf.de/infowiss/content/forschung/publikationen/MikeThelwall_DGD.pdf, Abruf: 2010-07-22.
- [TKVK+08] Tiago, P.; Kotilainen, N.; Vapa, M.; Kokkinen, H.; Nurminen, J.: Mobile Search - Social Network Search Using Mobile Devices. 2008, <http://kotilainen.eu/papers/MobileSearch.pdf>, Abruf: 2010-09-17.
- [Vaug08] Vaughan, L.: A New Frontier of Informetric and Webometric Research: Mining Web Usage Data. 2008, <http://www.collnet.de/Berlin-2008/VaughanWIS2008nfi.pdf>, Abruf: 2010-09-14.
- [WBHÖ+09] Weinhold, T.; Bekavac, B.; Hierl, S.; Öttl, S.; Herget, J.: Visualisierungen bei Internetsuchdiensten. In: Dirk Lewandowski (Hrsg.): Handbuch Internet-Suchmaschinen. Aka Verlag, Heidelberg, 2009, S. 249-282.
- [Weic09] Weichert, T.: Datenschutz bei Suchmaschinen. In: Dirk Lewandowski (Hrsg.): Handbuch Internet-Suchmaschinen. Aka Verlag, Heidelberg, 2009, S. 285-300.