

DONAU UNIVERSITÄT KREMS

Zentrum für praxisorientierte Informatik

Dr.-Karl-Dorrek-Str. 30

3504 Krems/Stein

Professional MSc in Informationstechnologien im Gesundheitswesen



Schlüsselfaktoren für glaubwürdige Gesundheitsinformationen: Web 2.0 oder Semantic Web?

Master Thesis

im Rahmen des Universitätslehrganges

Professional MSc in Informationstechnologien im Gesundheitswesen

verfasst und eingereicht von:

Johann Kalhs

Betreuer:

Dipl.-Ing. Christian H. Leeb

Erklärung an Eides statt

Ich, Johann Kalhs,
geboren am 16.4.1966 in Bad Aussee
erkläre,

dass ich meine Master Thesis selbstständig verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt und mich auch sonst keiner unerlaubten Hilfen bedient habe,

dass ich meine Master Thesis bisher weder im In- noch im Ausland in irgendeiner Form als Prüfungsarbeit vorgelegt habe,

dass ich, falls die Arbeit mein Unternehmen betrifft, meinen Arbeitgeber über Titel, Form und Inhalt der Master Thesis unterrichtet und sein Einverständnis eingeholt habe.

Ort, Datum

Unterschrift

Abstract

Every day, the World Wide Web is growing, offering and expanding information and knowledge. Accordingly, we are witnessing a rapid increase in users turning to the Web as a source of retrieving health information. At the same time, the reliability of such information has led to much debate among professionals, researchers, content providers and governmental authorities.

Collaborative knowledge systems has evolved in recent years and changed the view of the World Wide Web dramatically, leading to phenomena we now summarize under the phrase *Web 2.0*.

Now the question arises whether systems like *Wikipedia* can meet the standards that were set up for health information, because there is not much doubt that knowledge of that kind has been published as well as retrieved from *Wikipedia*.

When health seekers go online, they usually turn to Search Engines as a first resource of information. Collaborative knowledge systems like *Wikipedia* are ranked very prominently in search engines and at this stage provide the contents and the broad effect to present trustworthy health information.

This thesis shows that information taken from collaborative systems is valid and in many respects as reliable as peer-reviewed information.

In contrast, Semantic Web Applications have not become as widespread as anticipated in the past. From today's point of view, the way they are going to shape the future of information design can not be foreseen yet.

Further studies to validate the outcome of this work are recommended. Additionally to funding and supporting the actual development of software tools and technologies, it is strongly recommended to support providers of collaborative knowledge systems in order to improve the reliability of health information.

For this independent medium, which has led to a seismic shift in retrieving and generating global knowledge, contains the key to increasing the general awareness of healthcare issues.

Im World Wide Web erweitern sich Information und Wissen von Tag zu Tag. Dementsprechend lässt sich auch eine rapide Zunahme der Beschaffung von Gesundheitsinformationen durch Internetuser beobachten.

Zuverlässigkeit und Glaubwürdigkeit solcher Informationen haben zu Auseinandersetzungen zwischen Fachleuten, Wissenschaftlern, Anbietern solcher Dienste sowie Behörden geführt.

Kollaborative Wissenssysteme sind in den letzten Jahren entstanden und haben die Sicht auf das World Wide Web und dessen Nutzung dramatisch verändert. Diese und Ausprägungen wie Wikis, Blogs und Podcasts haben zu den Phänomenen geführt, die wir heute unter den Begriff Web 2.0 zusammenfassen.

Es besteht kein Zweifel daran, dass Gesundheitsinformationen in solchen Systemen vorhanden sind und genutzt werden. Können Systeme wie Wikipedia die Richtlinien für die Darstellung von Gesundheitsinformationen erfüllen?

Die Ergebnisse dieser Arbeit bestätigen, dass Wissen und Information aus diesen kollaborativen Systemen in vielerlei Hinsicht eine vergleichbare Glaubwürdigkeit hat wie solche aus expertengeprüften Systemen und dass Gesundheitsinformationen tatsächlich in solchen Systemen genutzt werden.

Suchmaschinen sind oft die erste Ressource, die Informationssuchende benutzen, um zu Gesundheitsinformationen zu gelangen. Kollaborative Wissensmanagementsysteme wie Wikipedia sind in Suchmaschinen sehr gut repräsentiert und bieten bereits heute die Inhalte und die nötige Breitenwirkung, um Gesundheitsinformationen für Nutzer glaubhaft darzustellen zu können.

Semantic Web Anwendungen sind noch nicht in großer Zahl vorhanden, ihr Einfluss auf die Informationsgestaltung lässt sich heute nicht zweifelsfrei festschreiben.

Weitere Studien zur Absicherung der Ergebnisse sind wünschenswert. Zusätzlich zur bereits laufenden Förderung von Forschung und Entwicklung von Web-Technologien zur Verbesserung der Qualität von Gesundheitsinformationen wird Behörden und Fördereinrichtungen die finanzielle Unterstützung von Anbietern von kollaborativen Wissensmanagementsystemen empfohlen.

Gerade auch in diesen unabhängigen Medien, die zu einer tektonischen Verschiebung der Wissensgenerierung führen, liegt ein Schlüssel zur Förderung und Erhöhung des Gesundheitsbewusstseins der Bevölkerung.

*We shall not cease from exploration
And the end of all our exploring
Will be to arrive where we started
And know the place for the first time*
T. S. Eliot

Danksagung

Diese Arbeit wäre ohne die gewinnbringenden Diskussionen mit meiner Kollegin und meinen Kollegen des Lehrgangs IT Healthcare sowie den Vortragenden auf Donau Universität Krems nicht möglich gewesen. Folgende Personen haben maßgeblich zum Abschluss dieser Arbeit beigetragen:

Herrn Dipl. Ing. Christian H. Leeb danke ich für Inspiration zum Thema Web 2.0 und die Überprüfung des Fortschritts während der Entstehungszeit dieser Arbeit.

Meine Frau, Dr. Tamara Kalhs, hat die medizinische Auswertung der Ergebnisse übernommen, dafür danke ich ihr herzlich.

Herrn Mag. Heinrich Tinhofer danke ich für die kritische Durchsicht der Arbeit vor der Drucklegung.

Herzlicher Dank gilt auch meinem Vorgesetzten, Dr. Roland Lavaulx-Vrécourt, Verwaltungsdirektor des St. Anna Kinderspitals. Er gewährte mir den nötigen Freiraum, dieses nicht unmittelbar mit meiner Tätigkeit verbundene Thema zu bearbeiten.

Inhaltsverzeichnis

Erklärung an Eides statt	II
Abstract.....	III
Danksagung	V
Inhaltsverzeichnis	VI
Verzeichnisse.....	IX
Abbildungsverzeichnis	IX
Tabellenverzeichnis.....	X
Stichwortverzeichnis	XI
Abkürzungsverzeichnis	XIII
1 Einführung und Hintergrund	15
1.1 Das World Wide Web	16
1.2 Geschichte der Qualitätsinitiativen für Gesundheitsinformationen	20
1.3 Das World Wide Web im Umbruch	22
1.4 Zentrale Fragestellungen.....	22
2 Material und Methoden	23
2.1 Vergleich Internet-Nutzungsstatistiken	23
2.2 Begriffsdefinitionen	24
2.3 Qualitätsinitiativen im World Wide Web.....	24
2.4 Literaturvergleich Wikipedia - Enc. Britannica	24
2.5 Aufbau und Durchführung der Feldanalyse	25
2.6 Praktische Anwendung eines <i>Code of Conduct</i>	26
3 Ergebnisse.....	27
3.1 Internet-Nutzungsstatistiken (Health-Seeker in USA und Europa)	27
3.1.1 Internet-Durchdringung.....	27
3.1.2 HIR - Health Information Retrieval	28
3.1.3 Benutzung von Suchmaschinen	31
3.1.4 Zufriedenheit der Health-Seeker mit der Gesundheitsinformation.....	32
3.2 Begriffserklärungen	32
3.2.1 Das Web 2.0	33
3.2.1.1 Internet Social Networking.....	34
3.2.1.2 User Generated Content (UGC).....	34
3.2.1.3 Harnessing Collective Intelligence (HCI)	35
3.2.2 Wikipedia	35

3.2.2.1	Das Revisionskontrollsystem der Wikipedia	37
3.2.2.2	Evolutionäre Artikelentwicklung (Selbsterhaltungskraft)	37
3.2.2.3	Vandalismus	37
3.2.2.4	Fälschung	37
3.2.3	Das Semantic Web	38
3.2.3.1	Metadaten	39
3.2.3.2	Repräsentation der Information (XML)	40
3.2.3.3	Repräsentation des Wissens (RDF)	40
3.2.3.4	Spezifizierung des Wissens (RDF Schema, OWL)	41
3.2.3.5	Ontologien	41
3.2.3.6	Exkurs: Ein Tag mit Michaels Semantic Web Agenten	42
3.3	Entwicklungsstand der Gesundheits-Qualitätsinitiativen	44
3.3.1	Codes of Conduct	44
3.3.2	Gütesiegelsysteme	45
3.3.3	Anleitungstools für Nutzer	45
3.3.4	Filtertechnologien	45
3.3.5	Zertifizierung durch Dritte - Third Party Rating	45
3.3.6	Medizinspezifische Suchmaschinen	46
3.4	Semantic Web Technologien bei Gesundheitsinformationen	47
3.4.1	UMLS	47
3.4.1.1	Metathesaurus	47
3.4.1.2	Semantisches Netzwerk	48
3.4.1.3	Specialist Lexicon	48
3.4.2	Dublin Core	48
3.4.3	MedCIRCLE	48
3.4.3.1	HIDDEL Vokabular für Metadaten	49
3.4.4	HIQuA	50
3.5	Wikipedia und Encyclopedia Britannica	52
3.6	Ergebnisse der Feldanalyse	54
3.6.1	Wikipedia Abfrage	54
3.6.2	Kontrollprobe	55
3.6.3	Google Abfrage	58
3.6.3.1	Ergebnisse der primären Krankheitsbegriffe	59
3.6.3.2	Ergebnis der Kontrollprobe	61
3.7	Ergebnisse der medizinischen Analyse	63
3.7.1	Umfang	64

3.7.2	Richtigkeit	65
3.7.3	Verständlichkeit für Laien.....	65
3.7.4	Gesamtbetrachtung der medizinischen Qualität	66
3.7.5	Kennzeichnungspraxis medizinischer Information in der Wikipedia.....	67
3.8	Zutreffen der Qualitätskriterien auf Wikipedia-Artikel	69
4	Diskussion	71
4.1	Health Information Retrieval	71
4.2	Wikipedia als Ressource für Gesundheitsinformationen.....	73
4.3	Schlussfolgerungen der medizinischen Analyse	76
4.4	Soziale Auswirkungen	77
4.5	Ausblick auf Semantic Web Anwendungen.....	77
4.6	Die Zukunft der klassischen Qualitätsinitiativen	79
4.7	Limitationen.....	80
5	Conclusio	82
5.1	Überprüfung der zentralen Fragestellungen	82
5.2	Ausblick	83
5.3	Empfehlungen	84
6	Anhang.....	85
6.1	Anteil der Internet-Zugänge von Haushalten der EU (%)	85
6.2	Internetaktivitäten - Einzelpersonen.....	86
6.2.1	Prozent der Einzelpersonen, Jahr 2006	86
6.2.2	Prozent der Internetnutzer, 2006.....	86
6.3	Gesundheitsrelevante Internetaktivitäten, 2006	87
6.3.1	Prozent der Einzelpersonen, 2006	87
6.3.2	Prozent der Internetnutzer, 2006.....	87
7	Literaturverzeichnis	88
7.1	Fachliteratur:	88
7.2	URI/URL	90

Verzeichnisse

Abbildungsverzeichnis

Abbildung 1: Internetbenutzer weltweit (Internet Worldstats)	19
Abbildung 2: Durchdringung der EU-Haushalte mit Internet (Eurostat, 2006)	28
Abbildung 3: Internet-Durchdringung in Europa (Eurostat)	29
Abbildung 4: Anteil der Health-Seeker in Europa (Eurostat, 2007)	30
Abbildung 5: Anteil der Health-Seeker in Österreich (Eurostat, 2007)	31
Abbildung 6: Semantic Web Layer (nach Koivunen / Miller 2001)	39
Abbildung 7: Design des HIQuA (nach Zeng et al, 2006).....	52
Abbildung 8: Entstehungszeit der Wikipedia-Artikel (Autor)	58
Abbildung 9: Verteilung des Rankings, primäre Krankheitsbegriffe (Autor)	59
Abbildung 10: Verteilung des Rankings, sekundäre Krankheitsbegriffe (Autor)	61
Abbildung 11: Verteilung von Umfang (Autor)	64
Abbildung 12: Verteilung von Richtigkeit (Autor)	65
Abbildung 13: Verteilung von Verständlichkeit (Autor)	66
Abbildung 14: Boxplots aller drei Wertungskriterien (Autor).....	67
Abbildung 15: Symbol zur Kennzeichnung von Gesundheitsartikel (Wikipedia)	68
Abbildung 16: Hinweis auf Gesundheitsinformation bei Wikipedia (Wikipedia)	68
Abbildung 17: The truth about the Semantic Web (nach Brickley, 2007)	78

Tabellenverzeichnis

Tabelle 1: Die 8 Kriterien des HONCode (nach Boyer 2005)	21
Tabelle 2: Ergebnisse der Wikipedia Abfrage (Autor).....	55
Tabelle 3: Ergebnisse der Wikipedia Kontrollprobe (Autor)	57
Tabelle 4: Google Rangliste der Primärabfrage (Autor)	60
Tabelle 5: Google Rangliste der Kontrollprobe (Autor).....	62
Tabelle 6: Ergebnisse der medizinischen Fachbewertung (Autor)	63
Tabelle 7: Zutreffen der Qualitätskriterien auf Wikipedia (Autor).....	70
Tabelle 8: Anteil der Internetzugänge in EU-Haushalten (Eurostat)	85
Tabelle 9: Anteil der Einzelpersonen in den letzten drei Monaten (Eurostat).....	86
Tabelle 10: Anteil der Internetnutzer in den letzten drei Monaten (Eurostat).....	86
Tabelle 11: Anteil der Health-Seeker (Eurostat)	87
Tabelle 12: Anteil der Health-Seeker in den letzten drei Monaten (Eurostat)	87

Stichwortverzeichnis

Begriff	Bedeutung
Collaboration ware	Kollaborative Software
Collaborative writing	Schriftwerk, von mehreren Personen gemeinschaftlich verfasst
community	Gruppe von Personen, die durch gemeinsames Wollen und Handeln verbunden ist
Digital Divide	Soziales Ungleichgewicht beim Zugang zur Informationsgesellschaft
Domain (1)	Domain im Sinne des Domain Name Systems
Domain (2)	Domain im Sinne einer Wissensdomain, etwa Medizin
Formale Semantik	Teilgebiet der Semantik, das sich mit der exakten Bedeutung von Sprachen (natürliche und künstliche) befasst. Besonders in der theoretischen Informatik, Logik und Linguistik wichtig
Health Information Retrieval	Die Nutzung von bzw. die Suche nach Gesundheitsinformationen im WWW
health seeker	Health-Seeker, in Amerika geprägte Bezeichnung für Konsumenten, die im Web nach Gesundheitsinformationen suchen
Knowledge Management	Wissensentwicklung, Wissenssammlung innerhalb einer Organisation, eines Unternehmens
Kollaborative Wissenssysteme	Wissenssammlungen erstellt durch kollaborative Arbeit, verteilte kollaborative Wissensmanagementsysteme, etwa <i>Wikipedia</i>
Lemma	Grundform eines Wortes; die Wortform, wie sie in einem Lexikon aufscheint
Mapping	Abbildung von Daten von einem Datenmodell zu einem unterschiedlichen anderen
network effect	Beschreibung des Effekts, dass der Nutzen eines Netzwerks umso größer wird, je mehr Nutzer daran teilnehmen (positive Rückkoppelung), z. B. Internet, WWW

Ontologie	In der Informatik: Datenmodell zur Repräsentation einer Sammlung von Konzepten sowie die Beschreibung des Zusammenhangs zwischen den Konzepten innerhalb einer Domain
peer collaboration	Werk, von mehreren Personen gemeinschaftlich verfasst
Semantic Web	Die Erweiterung des World Wide Web um maschinenlesbare Daten
Semantik	Bedeutungslehre in der Sprachwissenschaft, Bedeutung einer Informationsfolge in der Informatik
social bookmarking	Website für gemeinsame Verwaltung, Nutzung und Klassifizierung von Internet-Favoriten
Social Networking	Erscheinungsform der kollaborativen Zusammenarbeit verschiedener Interessensgruppen im Web
topic maps	ISO Standard für die Repräsentation und den Austausch von Wissen mit Schwerpunkt der leichteren Auffindbarkeit, Syntax XML-basiert, in vielerlei Hinsicht verwandt zu RDF
Vertikale Suchdienste	Suchmaschinen, die das Internet auf spezielle Themen oder Bereiche durchsuchen
wiki ware	Software für die Bereitstellung von Webinformationen mit sehr einfacher Syntax
wiki wiki	hawaiianisch für schnell, flott
Wissen	bewusste Information (eine von mehreren Definitionen)

Abkürzungsverzeichnis

CSS	Cascading Style Sheets
DCMI	Dublin Core Metadata Initiative
ESA	Explicit Semantic Analysis
Eurostat	Statistisches Amt der Europäischen Gemeinschaften
HCI	mehrere Bedeutungen, hier Harnessing collective intelligence
HIDDEL	Health Information Disclosure, Description and Evaluation Language
HIR	Health Information Retrieval
HON	Health on the Net Foundation
IDC-10	WHO: International Statistical Classification of Diseases and Related Health Problems, 10. Revision
KM	Knowledge Management
LOINC	Logical Observation Identifiers Names and Codes
MedPICS	Medical Platform for Internet Content Selection, ein auf PICS basierendes Metadaten-set für Health Information
MeSH	Medical Subject Headings
MIT	Massachusetts Institute of Technology
MRCOC	UMLS co-occurrence table
MRREL	UMLS metathesaurus relationship table
NLM	National Library of Medicine
OWL	Web Ontology Language
PHP	Hypertext Preprocessor, Skriptsprache zu Erstellung dynamischer Webinhalte
PICS	Platform for Internet Content Selection, ein früher W3C-Standard zur Erhöhung der Interoperabilität im WWW
RDF	Resource Description Framework
SGML	Standard Generalized Markup Language
SNOMED	Systematized Nomenclature of Medicine
UGC	user generated content
UMLS	Unified Medical Language System
URAC	Utilization Review Accreditation Commission
URI	Universal Resource Identifier, ursprüngliche Bezeichnung für URL

URL	Uniform Resource Locator
XML	Extensible Markup Language

1 Einführung und Hintergrund

Die Verbreitung und auch die Nachfrage von Information und Wissen im Internet gehen mit ungebrochener Geschwindigkeit voran. Das World Wide Web ist als Medium zu Informationsbeschaffung nicht mehr wegzudenken.

Eine immer größer werdende Zahl von Menschen benutzt das Internet, um Wissen, speziell auch über gesundheitsrelevante Themen zu erlangen. Die Abfrage von Gesundheitsinformationen ist für Millionen von Menschen eine wichtige und übliche Gewohnheit geworden.

Ein großes Hemmnis für die Akzeptanz ist nach wie vor die Sorge über die Qualität, Glaubwürdigkeit und Zuverlässigkeit solcher Informationen. Sind die Informationsanbieter vertrauenswürdig? Sind die Informationen richtig? Das sind wichtige Fragen, die sich Information suchende Nutzer stellen¹.

In diesem Spannungsfeld zwischen der Quantität und der Qualität dieser Informationen haben sich ab dem letzten Jahrzehnt des 20. Jahrhunderts weltweit verschiedene Qualitätsinitiativen entwickelt, deren Ziel es ist, glaubwürdige Informationen gegenüber ungläubwürdigen oder gar betrügerischen zu kennzeichnen. Diesen Initiativen liegen ethische Standards zugrunde, die in so genannten *Codes of Conduct* festgeschrieben werden.

Es steht also seit mehr als einem Jahrzehnt eine Vielzahl von Bewertungstools für Informationssuchende zur Verfügung. Viele dieser Tools sind jedoch für Laien in der Anwendung zu kompliziert oder schlicht ungeeignet.

Außer den einfachen Gütesiegelsystemen haben sich kaum Systeme mit aufwändigeren Technologien etablieren können. Regulierungsversuche mit Metadatenmodellen und Semantic Web Modellen, auch von Behörden unterstützt, sind bisher nicht über den Status eines Experiments hinausgekommen.

Die Nachfrage nach Gesundheitsinformationen im World Wide Web ist aber ungebrochen hoch, seit dem Jahr 2006 hat sich in der englischsprachigen Fachliteratur dafür bereits

¹ Vgl. Risk, Ahmad / Joan Dzenowagis, 'Review of internet health information quality initiatives' (2001) 3(4) *J Med Internet Res* E28.

ein eigener Fachbegriff, *Health Information Retrieval*, etabliert², und ein nach Gesundheitsinformation suchender Nutzer wird als *Health-Seeker*³ bezeichnet.

Seit der Jahrtausendwende haben sich kollaborative Wissenssysteme entwickelt, die heute zu weit verbreiteten und akzeptierten Informationsmedien zählen.⁴

Mit der als *Web 2.0* bezeichneten Sicht auf das World Wide Web stehen nun Systeme zur Verfügung, die große Reichweiten gewinnen. Diese kollaborativen Systeme bauen auf der Wissensbasis breiter Bevölkerungsschichten auf. Sie bilden ein geeignetes Fundament, auf dem auch Gesundheitsinformationen zuverlässig dargestellt werden können. Den Anforderungen an die Ethik können solche Systeme ebenso gerecht werden wie autorenbasierte Systeme.

Sind diese Systeme - hier im Speziellen *Wikipedia*⁵ - geeignet, um Gesundheitsinformationen glaubwürdig und vertrauenswürdig darzustellen? Bringen sie Nutzen für den Patienten, seine Angehörigen und Freunde, das sind die zentralen Fragen, die in dieser Arbeit behandelt werden.

1.1 Das World Wide Web

Das World Wide Web durchdringt immer mehr Lebensbereiche des Menschen. Die Möglichkeiten, wie Menschen miteinander kommunizieren, wie Informationen verbreitet und gewonnen werden und wie Geschäfte durchgeführt werden, haben sich dramatisch verändert.

Die 1989 von Timothy Berners-Lee propagierte Idee des World Wide Webs als ein freier Informationsraum im Internet ist heute zur zentralen Plattform in der Online-Welt geworden. Das World Wide Web ist das größte Sammelbecken für Information und Wissen jeglicher Art.

Begriffe wie *URL (URI⁶)*, *HTTP* und *HTML*, nach abnehmender Bedeutung sortiert, beschreiben die wenigen Basisregeln, die das World Wide Web ausmachen und die zu

² Vgl. Zeng, Q. T. et al, 'Assisting consumer health information retrieval with query recommendations' (2006) 13(1) *J Am Med Inform Assoc* 80-90.

³ Vgl. Fox, Susanna *Online Health Search 2006* (2006)
<http://www.pewinternet.org/pdfs/PIP_Online_Health_2006.pdf> am 03.03.2007, S. 1.

⁴ Vgl. Boulos, M. N. Kamel / I. Maramba / S. Wheeler, 'Wikis, blogs and podcasts: a new generation of Web-based tools for virtual collaborative clinical practice and education' (2006) 6 *BMC Med Educ* 41.

⁵ *Wikipedia* ist die weltweit größte und in vielen Sprachen verfügbare Online-Enzyklopädie.

⁶ URI, Universal Resource Identifier, von Berners-Lee favorisierter Begriff für URL.

einer noch nie da gewesenen Granularität und Vielfalt an Information geführt haben. Diese Begriffe sind heute jedem Internet-Anwender geläufig.

Die Vision des Gründers war es, ein System zu schaffen, in dem das Beitragen des eigenen Wissens genau so einfach sein sollte wie das Abrufen dessen, was jemand anderer publiziert hat.⁷

Von Beginn an wollte Berners-Lee, dass eine Software-Anwendung für das World Wide Web - was später als Web-Browser bzw. Web-Editor bekannt wurde - gleichermaßen sowohl für die Betrachtung der Information als auch zum Editieren der Information geeignet war. Die Vision des kollaborativen Mediums ist also so alt wie das Web selbst.

Die Entwickler der Web-Browser, von 1992 bis zum heutigen Tag, folgen dieser Idee jedoch nicht wirklich. Es zeigt sich, dass diese Vision mehr als ein Jahrzehnt brauchte, um - hauptsächlich durch die die Kraft der *Community*⁸ - verwirklicht zu werden.

Das fundamentale Prinzip, das dem World Wide Web zugrunde liegt, bringt Berners-Lee wie folgt zum Ausdruck:

“...once someone somewhere made a document, database, graphic, sound, video or screen at some stage in an interactive dialogue, it should be accessible (subject to authorisation, of course) by anyone, with any type of computer, in any country. And it should be possible to make a reference - a link - to that thing, so that others could find it”.⁹

Als 1993 verschiedene Institutionen und Unternehmungen begannen, den Nutzen des World Wide Web zu erkennen, war die Gefahr gegeben, dass das Web in verschiedene Bereiche, akademische, kommerzielle, freie sowie kostenpflichtige, zerfallen könnte. Die Aktivitäten verlagerten sich vom europäischen CERN¹⁰, an dem Berners-Lee das Web entwickelte, an das amerikanische MIT¹¹, das größeres Augenmerk auf die Entwicklung von praktischen Anwendungen für das Web legt.

Nachdem das CERN auf sein Lizenzrecht verzichtet hatte, war der Weg frei, das World Wide Web mit wenigen Basisregeln zu standardisieren, um so die Gefahr des Zerfalls

⁷ Vgl. Berners-Lee, Tim / Mark Fischetti, *Weaving the Web : The Past, Present and Future of the World Wide Web by its inventor* (1999), S. 36.

⁸ *Community* ist der Begriff für eine Gruppe von Personen, die gemeinsames Wollen und Wissen entwickelt.

⁹ Berners-Lee, Tim / Mark Fischetti, *Weaving the Web : The Past, Present and Future of the World Wide Web by its inventor* (1999), S. 40.

¹⁰ Europäische Organisation für Kernforschung, ursprünglich *Conseil Européen pour la Recherche Nucléaire*.

¹¹ Massachusetts Institute of Technology (Cambridge, USA).

hintan zu halten. Das W3C¹² wurde gegründet, dessen erster und heutiger Direktor Tim Berners-Lee ist.

In den späten 90er Jahren des vergangenen Jahrtausends setzte die starke Kommerzialisierung des World Wide Web ein. Hinlänglich bekannt ist vor allem auch der Zusammenbruch der Börsenkurse der meisten Internet-Unternehmen, deren Kurse enorm überbewertet waren. Dieses spektakuläre Phänomen ist auch als Platzen der Dotcom-Blase bekannt geworden.

Das Vertrauen in die börsennotierten Internet-Unternehmen ist dadurch merklich zurückgegangen, das in das World Wide Web aber nicht.

So hat sich das Web in technologischer Hinsicht und auch hinsichtlich der Nutzungsmöglichkeiten weiter entwickelt, auch die Sicht bzw. die Wahrnehmung auf das Web veränderte sich wiederum.

Viele Anwendungen, die früher mit Einzelapplikationen auf einem lokalen Computer durchgeführt worden sind, werden heute über das World Wide Web erledigt, etwa Verwaltung und gemeinsame Nutzung von Photos (flickr.com) oder Verwaltung von Hyperlinks (del.icio.us¹³).

Heute sieht man einen Computer als den Einstiegspunkt zum Information Highway, während man noch vor kurzer Zeit eher ein Gerät zur Kalkulation und Textverarbeitung damit assoziiert hat¹⁴.

Die Zahl der Benutzer, die Informationen aus dem World Wide Web beziehen bzw. die das Internet zur Kommunikation benutzen, ist stetig im steigen. Die aktuellen Zahlen belegen, dass im März 2007 bereits 1,1 Milliarden Menschen Zugang zum Internet hatten. Diese Anzahl entspricht einem Weltbevölkerungsanteil von ca. 17% bei einer für 2007 geschätzten Weltbevölkerung von 6,6 Milliarden Menschen (Abb. 1).

¹² World Wide Web Consortium.

¹³ Kollaborative Website zur Verwaltung von Hyperlinks (auch „Social Bookmarking“).

¹⁴ Vgl. Antoniou, G. / Frank van Harmelen, *A semantic Web primer*, Cooperative information systems (2004).

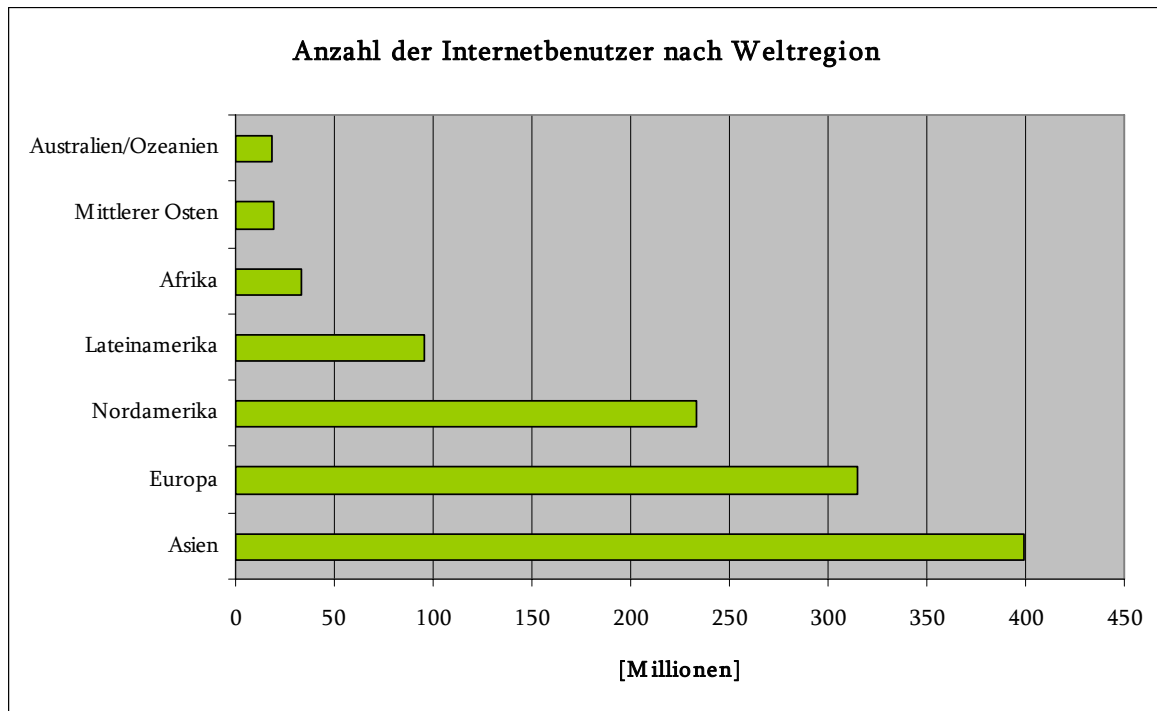


Abbildung 1: Internetbenutzer weltweit (Internet Worldstats¹⁵)

Die Durchdringungsraten in den einzelnen Kontinenten sind dabei sehr unterschiedlich. In Nordamerika liegen sie bei 69,7%, Australien und Ozeanien bei 53,5% und in Europa bei 38,9%. Asien hat bei einem Weltbevölkerungsanteil von mehr als 56% eine Internet-Durchdringungsrate von 10,7%. Afrika hat mit 3,5% die niedrigste Durchdringungsrate aller Kontinente.¹⁶

Während in den klassischen Industrieländern die Zuwachsraten sinken bzw. auf hohem Niveau stagnieren, ist in den bevölkerungsreichen Kontinenten Asien und Afrika ein hoher Anstieg der Zuwachsraten zu beobachten. Eben durch diese Kontinente ist mit einem weiteren starken Steigen der Benutzerzahlen zu rechnen. Damit ist offensichtlich, dass das World Wide Web seine Stellung als global wichtiges Medium weiter ausbauen wird.

Ob dadurch auch das als Digital Divide bezeichnete soziale Ungleichgewicht ausgeglichen werden kann, ist fraglich. Es ist immer noch so, dass vor allem den

¹⁵ o.V., *World Internet Users and Population Stats* (2007) Miniwatts Marketing Group
<<http://www.internetworldstats.com/stats.htm>> am 18.04.2007 .

¹⁶ o.V., *World Internet Users and Population Stats* (2007) Miniwatts Marketing Group
<<http://www.internetworldstats.com/stats.htm>> am 18.04.2007 .

gebildeten und sozial besser gestellten Gruppen der Weltbevölkerung der Zugang wesentlich leichter möglich ist, als denjenigen, die nicht so gut ausgebildet sind¹⁷.

Gerade aber der Zugang zu Gesundheitsinformationen, im Sinne von „gesund werden“, „gesund bleiben“, ist auch von großer Bedeutung für die finanzielle und soziokulturelle Entwicklung der Volkswirtschaften.

1.2 Geschichte der Qualitätsinitiativen für Gesundheitsinformationen

Die Geschichte der Qualitätskontrolle und Kennzeichnung von Gesundheitsinformationen im World Wide Web geht bis zum Jahr 1994, also bis zu dem Zeitpunkt zurück, ab dem das World Wide Web in den Fokus kommerzieller Interessen gerückt ist.

Aufbauend auf Ethikstandards, die von Medizinern, Wissenschaftlern und Ethikern in mehreren Ländern unabhängig voneinander formuliert worden sind, leiten sich Verhaltenskodizes, die sehr allgemein als *Codes of Conduct* bezeichnet werden, ab. Diese sind ihrerseits wiederum die Grundlage für technologiebasierte Lösungen zur Überprüfung der Glaubwürdigkeit von Gesundheitsinformationen im World Wide Web. Mit den Verhaltenskodizes wurden zwei Ziele verfolgt. Einerseits wurde versucht, die Nutzer von Gesundheitsinformationen vor Fehlinformation und Betrug zu schützen. Andererseits wurden den Anbietern im World Wide Web Richtlinien für die Darstellung ihrer Gesundheitsinformationen empfohlen. Diese Verhaltenskodizes hatten keinerlei rechtliche Verbindlichkeit und etablierten sich in verschiedenen Ländern unabhängig voneinander.

Exemplarisch ist der *Code of Conduct* der HON¹⁸, einer der wichtigsten Vertreter von Gütesiegelsystemen. Er wurde 1996 festgeschrieben und ist mittlerweile in 29 Sprachen übersetzt¹⁹.

Die wesentlichen Kriterien, die sich auf die gesundheitsrelevanten Inhalte anwenden lassen sollen sind (1) Urheberschaft, (2) Ergänzung, (3) Vertraulichkeit, (4) Quellenangabe, (5) Wissenschaftliche Begründung, (6) Erreichbarkeit der Autoren, (7) Transparenz der Sponsoren, (8) Ehrlichkeit bei Werbung und redaktioneller Politik. Tabelle 1 gibt die im HON *Code of Conduct* verwendeten Begriffe und deren Kurzbeschreibung wieder.

¹⁷ Vgl. Fox, Susanna *Online Health Search 2006* (2006)

<http://www.pewinternet.org/pdfs/PIP_Online_Health_2006.pdf> am 03.03.2007, S.3.

¹⁸ HON = Health on the Net Foundation, <http://www.hon.ch>.

¹⁹ Vgl. Boyer, C. / A. Geissbuhler, 'A decade devoted to improving online health information quality' (2005) 116 *Stud Health Technol Inform* 891-896, S. 892.

Kriterium	Beschreibung
Authority	Wird medizinische Hilfe durch einen Mediziner gegeben?
Complementarity	Information muss unterstützen, nicht die Arzt-Patientenbeziehung ersetzen
Privacy	Wie werden personenbezogene Daten - die durch eine Website gesammelt werden können - verwendet?
Attribution	Gewissenhafte und genaue Referenzierung auf die Informationsquellen
Justifiability	Die Wirksamkeit und der Nutzen eine Therapie oder eines Produktes werden ausgewogen dargestellt
Transparency	gut zugänglich Darstellung der Identität von Editoren und Verantwortlichen, Kontaktmöglichkeit per email
Financial disclosure	Klare Benennung aller Sponsoren, die finanziell zu Entstehung und zum Betrieb beitragen
Sponsorship	Klare Trennung des redaktionellen Inhalts von Werbeinhalt

Tabelle 1: Die 8 Kriterien des HONCode (nach Boyer 2005)

Die Mission wird von HON formuliert als „*guide the growing online community of healthcare consumers and information providers to sound, reliable medical information and expertise*“²⁰.

Die meisten Initiativen haben sich vor der Jahrtausendwende etabliert, einige sind mit unveränderter Technologie heute noch aktiv, andere haben versucht durch fortgeschrittene Technologien ihren Nutzwert zu erhöhen.

Ein detaillierter Überblick wurde zuletzt von der WHO²¹ gegeben. Weitere Sichtungen der Literatur von 2001 bis heute haben nur punktuell Neuerungen ergeben, keine Initiative konnte wesentlich an Breitenwirkung zulegen. Das Interesse an Forschung und Entwicklung in diesem Bereich scheint heute weniger ausgeprägt zu sein.

Allen Initiativen, die auf Verhaltenskodizes basieren ist gemeinsam, dass sie hauptsächlich die formalen Kriterien der dargebotenen Gesundheitsinformation (in Form von Herkunft, Präsentation, Aktualität) bewerten, und nicht die medizinische

²⁰ Boyer, C. / A. Geissbuhler, 'A decade devoted to improving online health information quality' (2005) 116 *Stud Health Technol Inform* 891-896, S. 891.

²¹ Vgl. Risk, Ahmad / Joan Dzenowagis, 'Review of internet health information quality initiatives' (2001) 3(4) *J Med Internet Res* E28.

Information an sich. Dies beruht auf der Tatsache, dass das medizinische Angebot im Web so umfangreich ist und sich ständig erweitert, dass eine Beurteilung auch durch Fachpersonal mit vertretbarem Aufwand nicht mehr durchführbar ist.

1.3 Das World Wide Web im Umbruch

Mit der Verbreitung von Technologien und Praktiken, die heute mit dem Begriff *Web 2.0* in Verbindung gebracht werden, ist eine dramatische Veränderung des Nutzungsverhaltens des World Wide Web sichtbar geworden.

Besonders mit kollaborativen Wissenssystemen, hier wiederum besonders die Online-Enzyklopädie Wikipedia, und mit Systemen, die von einer großen Community getragen werden, haben sich die Informationsgestaltung und die Informationsverbreitung stark gewandelt.

Während man als *Semantic Web* die Weiterentwicklung des World Wide Webs um maschinenlesbare Daten bezeichnet, also durchaus technologiebezogen ist, ist der Begriff *Web 2.0* ein unscharfer Oberbegriff für die heutige Wahrnehmung des World Wide Webs, der stark von interaktiven, kollaborativen Systemen mit Millionen von teilnehmenden Nutzern geprägt ist.^{22 23}

1.4 Zentrale Fragestellungen

Sind kollaborative Systeme geeignet, Gesundheitsinformation darzustellen?

Die kollaborativen Wissenssysteme im World Wide Web haben sich heute zu ernst zu nehmenden Informationsquellen für Health-Seeker entwickelt. Können solche Systeme, besonders die Online-Enzyklopädie Wikipedia, Gesundheitsinformationen glaubwürdig darstellen?

Ist Wikipedia die erste Ressource solcher Gesundheitsinformationen, bei der auch die medizinischen Inhalte selbst hohe Glaubwürdigkeit haben? Welchen Einfluss werden die aufkeimenden *Semantic Web* Applikationen auf die Darstellung und auf die Abfrage von Gesundheitsinformationen haben? Das sind die zentralen Fragen, die in dieser Arbeit beleuchtet werden.

²² Vgl. Kaplan, Craig A., 'Collective Intelligence: a new approach to stock price forecasting' (Arbeit präsentiert bei den Proceedings of the 2001 IEEE Systems, Man, and Cybernetics Conference, Tucson, AZ, USA, 2001).

²³ O'Reilly, Tim, *Web 2.0 Compact Definition: Trying Again* (2006) O'Reilly Radar
<http://radar.oreilly.com/archives/2006/12/web_20_compact.html> am 02.03.2007 .

2 Material und Methoden

Die Eignung kollaborativer Wissenssysteme für die Bereitstellung von glaubwürdigen Gesundheitsinformationen wird in dieser Arbeit einerseits mit einem Literaturvergleich und andererseits mit einer Feldanalyse im World Wide Web untersucht.

Die aktuellen Durchdringungsraten und die Verbreitung von Internetzugängen in den USA und in Europa, bzw. in einigen ausgewählten Europäischen Ländern, werden gesammelt und vergleichend dargestellt.

Das Internetnutzungsverhalten bei der Beschaffung nach Gesundheitsinformationen aus den USA und Europa wird verglichen und den Ergebnissen der Analyse gegenübergestellt. Der aktuelle Stand der Qualitätsinitiativen bzw. der Technologien zur Überprüfung der Qualität von Gesundheitsinformationen im World Wide Web wird dargestellt.

Ebenso werden Studien über kollaborativen Wissensmanagementsysteme und semantische Webtechnologien verglichen und den Ergebnissen der Analyse gegenübergestellt.

Bei der Feldanalyse werden Begriffe von verbreiteten Krankheiten in der deutschen Wikipedia und auf der Suchmaschine Google abgefragt und analysiert. Der Informationsgehalt wurde von einem Mediziner auf Umfang, Glaubwürdigkeit, Eignung für Laien geprüft.

Schließlich wird die Frage geklärt, ob die in der Literatur genannten Qualitätskriterien für Gesundheitsinformationen im World Wide Web in Wikipedia eingehalten werden können.

Das Internet ist mehr als das World Wide Web, in dieser Arbeit werden die Begriffe dennoch synonym verwendet.

2.1 Vergleich Internet-Nutzungsstatistiken

In der Literatur werden Ergebnisse von Studien über das Nutzungsverhalten und die demographische Verteilung der Nutzer im World Wide Web gesucht und vergleichend dargestellt.

Ebenso werden detaillierte Studien über Nutzerstatistiken bei der Abfrage von Gesundheitsinformationen, die in dieser spezifischen Form nur in den USA zur Verfügung stehen, herangezogen.

In Europa gibt es wenige Studien mit vergleichbarem Detaillierungsgrad. Hier wurden punktuell Daten aus einschlägigen Studien von einigen Mitgliedsländern des EWR (Norwegen, Großbritannien) bzw. Daten, die vom Statistischen Amt der Kommission der

Europäischen Union²⁴ im Internet bereitgestellt werden, analysiert und vergleichend dargestellt.

2.2 Begriffsdefinitionen

Zentrale Begriffe, deren Bedeutungen für das Verständnis dieser Arbeit unbedingt erforderlich sind und die die heutige Entwicklung des World Wide Web bezeichnen, werden in einem eigenen Unterkapitel geklärt. Dabei handelt es sich um eine Zusammenschau aus der verfügbaren Literatur bezüglich der Begriffe Web 2.0, Semantic Web und dem Phänomen Wikipedia bzw. um Begriffe und Standards, die mit diesen im direkten Zusammenhang stehen. Diese Definitionen sind in der weiteren Folge für die Betrachtung der Bewertungstools der Qualitätsinitiativen von Bedeutung.

2.3 Qualitätsinitiativen im World Wide Web

Ausgehend von der letzten in der Literatur gegebenen Übersicht der Qualitätsinitiativen²⁵ wird geklärt, welche neuen Entwicklungen seit 2001 stattgefunden haben und welche Anwendungen dem Benutzer heute zur Evaluierung der Qualität von Gesundheitsinformationen zur Verfügung stehen.

Wo steht die Entwicklung bei Semantic Web Technologien bezüglich Health Information Retrieval, sind bereits Anwendungen verfügbar, das ist eine weitere Frage, die in der aktuellen Fachliteratur recherchiert wird.

2.4 Literaturvergleich Wikipedia - Enc. Britannica

Hinsichtlich des Gebrauchs von Wikipedia für die Beschaffung von Gesundheitsinformationen liegen noch keine bzw. sehr wenige aktuelle Untersuchungen vor. Die Untersuchung, inwieweit Wikipedia ein für die Darstellung solcher Informationen geeignetes Medium darstellt, ist ein Schwerpunkt dieser Arbeit.

Allgemeine verfügbare Arbeiten und Studien über Wikipedia hinsichtlich Technologie, Glaubwürdigkeit und Verbreitung werden dargestellt.

²⁴ Abkürzung Eurostat.

²⁵ Vgl. Risk, Ahmad / Joan Dzenowagis, 'Review of internet health information quality initiatives' (2001) 3(4) *J Med Internet Res* E28

Es liegt eine vergleichende Studie über die Qualität von Artikeln in Online-Enzyklopädien, zwischen der Wikipedia und der Encyclopedia Britannica²⁶, vor. Diese Studie wird zur Klärung der These herangezogen.

2.5 Aufbau und Durchführung der Feldanalyse

Mithilfe eines Mediziners wurde eine Liste von 60 Krankheiten bzw. Diagnosen mit den in Österreich dafür allgemein gebräuchlichen, deutschsprachigen oder lateinischen, Bezeichnungen erstellt.

Aus dieser Liste von Krankheiten, so genannter Volkskrankheiten, wird nach dem Zufallsprinzip eine Stichprobe (N=30) ausgewählt und einer näheren Untersuchung im World Wide Web zugeführt.

Ein jeweiliger Begriff aus dieser Liste wird auf der Website der deutschen Wikipedia²⁷ abgefragt. Danach wird jeder dieser Begriffe auch auf den Suchdienst Google Österreich²⁸ (Suchmethode „das Web“) angewendet.

Besonderer Wert wird darauf gelegt, allgemeine und einfache Suchbegriffe zu formulieren, so wie sie auch ein Nutzer bei der Suche nach Gesundheitsinformationen üblicherweise verwendet²⁹ ³⁰. Es werden daher die weit verbreiteten deutschen oder lateinischen Bezeichnungen der Krankheiten ohne zusätzliche Angaben abgefragt.

Aus den Ergebnissen der Suchmaschinen-Abfrage (n=30) wird analysiert, welche Platzierung (PageRank³¹) die Artikel aus der deutschen Wikipedia erreichen, und welchen anderen Informationsanbieter die Ergebnisliste als Erstgereihten anführt, so es nicht Wikipedia ist.

Im Allgemeinen ist davon auszugehen, dass prominent gereichte Einträge von den Informationssuchenden bevorzugt weiterverfolgt werden.

Die Ergebnisse der Abfragen in Wikipedia (N=30) wurden hinsichtlich statistischer Daten (Vorhandensein, Anzahl der Versionen, Erstellungsdatum, Umfang) analysiert.

²⁶Vgl. Giles, Jim, 'Internet encyclopaedias go head to head' (2005) 438 *Nature* 900-901.

²⁷ <http://de.wikipedia.org>.

²⁸ <http://www.google.at>.

²⁹ Vgl. Sim, Nurain Z. et al, 'Information on the World Wide Web—how useful is it for parents?' (2007) 42 *Journal of Pediatric Surgery* 305-312, S. 306.

³⁰ Vgl. Zeng, Q. T. et al, 'Assisting consumer health information retrieval with query recommendations' (2006) 13(1) *J Am Med Inform Assoc* 80-90, S80.

³¹ PageRang, Technologie zur Gewichtung der Suchergebnisse bei Google.

Die Gesundheitsinformation der einzelnen Artikel an sich, der medizinische Inhalt des jeweiligen Artikels in Wikipedia wird von einem Facharzt in Bezug auf die medizinischen Qualität, des Umfangs und der Eignung für den Laien beurteilt. Die Beschränkung auf einen Mediziner ist aus praktischen Gründen notwendig. Der Aufwand, mehrere sachkundige Laien selbst zur Beurteilung heranzuziehen, ist im Rahmen der Erstellung der Masterthese nicht möglich.

Um systematische Fehler, die aufgrund der geringen Stichprobenzahl von 30 auftreten könnten, zu vermeiden, wird nach der Durchführung der Abfrage und der nachfolgenden Analyse der primären Gruppe ebenso die Gruppe der verbliebenen 30 Krankheiten in Wikipedia und Google abgefragt und nach den beschriebenen Methoden analysiert. Die zweite Gruppe wird jedoch nicht mehr auf medizinische Relevanz geprüft.

Sofern im geringen Umfang statistische Auswertungen erforderlich sind, werden diese mit SPSS Version 13.0 durchgeführt. Für die Auswertung der Liste der Versionsentwicklung bei Wikipedia, die jedem Artikel zugeordnet ist, wird der frei erhältliche Editor *Weaverslave* verwendet.

2.6 Praktische Anwendung eines *Code of Conduct*

Es wird untersucht, inwieweit die in Wikipedia dargestellten Gesundheitsinformationen den wesentlichen Kriterien der in der Literatur genannten *Codes of Conduct* für Gesundheitsinformation im World Wide Web entsprechen.

3 Ergebnisse

3.1 Internet-Nutzungsstatistiken (Health-Seeker in USA und Europa)

Gesundheitsinformationen, Healthcare Information, haben seit Beginn des Webzeitalters einen großen Stellenwert. Fachleute und Forscher tauschen Informationen aus, Lehre wird teilweise über das World Wide Web betrieben.

Die Gruppe der Benutzer (Laien, Betroffene oder deren Angehörige), englisch Health-Seeker, steht im Mittelpunkt dieser Arbeit.

3.1.1 Internet-Durchdringung

Die Durchdringung der Bevölkerung mit Internet ist weltweit sehr unterschiedlich ausgeprägt³². Sie ist in den USA etwas höher als in Europa. Die Zuwachsraten sind in beiden Ländern vergleichbar^{33 34 35}. Auch die Verteilung der Altersgruppen ergibt ein ähnliches Bild.

Abbildung 2 zeigt die Durchdringung der Haushalte mit Internetanschlüssen im Jahr 2003 sowie im Jahr 2006 für verschiedene Europäische Länder und gesamt für die Länder der Europäischen Union (15 Mitgliedsländer).

Sim et al. berichten in der 2007 erschienen Studie über die Ergebnisse einer Befragungen der Patienten des Great Ormond Street Hospital in London über die Nützlichkeit des Internet für Patienten, dass der Anteil von Internetanschlüssen in britischen Haushalten von 9% (1998-1999) auf 53% in 2004-2005 gestiegen ist. Die Daten von Eurostat³⁶ zeigen für das Jahr 2006 einen Wert knapp über 60%. Der Anteil von Nutzern, die das Web für Gesundheitsinformationen benutzten liegt in Großbritannien unter dem Wert der USA³⁷.

³² Vgl. Kap. 1.1, S. 16.

³³ Vgl. Fox, Susanna *Online Health Search 2006* (2006)

<http://www.pewinternet.org/pdfs/PIP_Online_Health_2006.pdf> am 03.03.2007

³⁴ Vgl. Fox, Susanna / Lee Rainie, *Vital decisions* (2002)

<http://207.21.232.103/pdfs/PIP_Vital_Decisions_May2002.pdf> am 16.04.2007

³⁵ Vgl. o.V., *Internet Monitor Austria - Kommunitaktion und IT in Österreich* (2006)

<http://www.integral.co.at/dlimages/AIM-C_4Quartal_2006.pdf> am 03.03.2007 .

³⁶ Vgl. o.V., *Eurostat Leitseite* Europäische Kommission

<http://epp.eurostat.ec.europa.eu/portal/page?_pageid=1090,30070682,1090_33076576&_dad=portal&_schema=PORTAL> am 11.03.2007 .

³⁷ Vgl. Sim, Nurain Z. et al, 'Information on the World Wide Web—how useful is it for parents?' (2007) 42 *Journal of Pediatric Surgery* 305-312, S. 306.

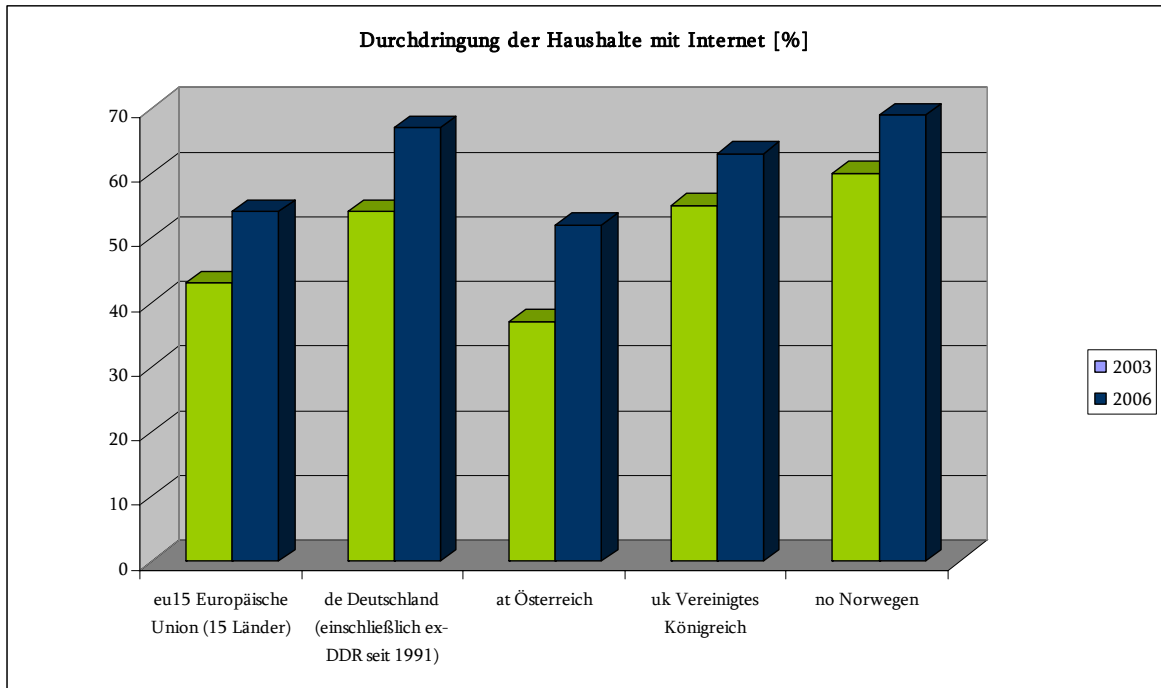


Abbildung 2: Durchdringung der EU-Haushalte mit Internet (Eurostat, 2006)

3.1.2 HIR - Health Information Retrieval

Abfragen nach Gesundheitsinformationen im World Wide Web sind bereits so etabliert, dass in der Fachliteratur der Begriff *HIR* verwendet wird, der für Health Information Retrieval steht^{38 39}.

Im Jahr 2000 lag die Zahl der Internetnutzer in den USA, die zumindest einmal pro Jahr nach Gesundheitsinformation suchten, bei 52 Millionen Menschen. Im Jahr 2002 stieg diese Zahl auf 73 Millionen bzw. 62% aller Internetnutzer. Im Jahr 2006 suchten 80% aller Internetnutzer mindestens ein Mal pro Jahr nach Gesundheitsinformationen, zur Hälfte für sich selbst, zur anderen Hälfte aber für Freunde oder Verwandte.⁴⁰

Der typische Nutzer von Gesundheitsinformationen ist jünger als 65 und gut ausgebildet. 85% aller 18-29 Jahre alten amerikanischen Bürger haben Zugang zum Internet. Diese Rate sinkt mit zunehmendem Alter, liegt bei der Gruppe der 50-64 Jährigen bei 69% um bei der Gruppe der 65+ Jährigen auf 31% abzusinken.

³⁸ Vgl. Hesse, Bradford W et al, 'Trust and Sources of Health Information. The Impact of the Internet and Its Implications for Health Care Providers: Findings From the First Health Information National Trends Survey' (2005) 165(22) *Arch Intern Med* 2618-2624

³⁹ Vgl. Zeng, Q. T. et al, 'Assisting consumer health information retrieval with query recommendations' (2006) 13(1) *J Am Med Inform Assoc* 80-90

⁴⁰ Vgl. Fox, Susanna / Lee Rainie, *Vital decisions* (2002)

<http://207.21.232.103/pdfs/PIP_Vital_Decisions_May2002.pdf> am 16.04.2007

2006 hatten Menschen aus der Gruppe mit Grundausbildung (High school diploma) einen Anteil von 59%, die mit einer höheren Ausbildung (college degree) von 91%. Die Schere zwischen mittlerer Ausbildung und höherer Ausbildung ist vom Jahr 2000 (39%) bis 2006 auf 32% gesunken.⁴¹

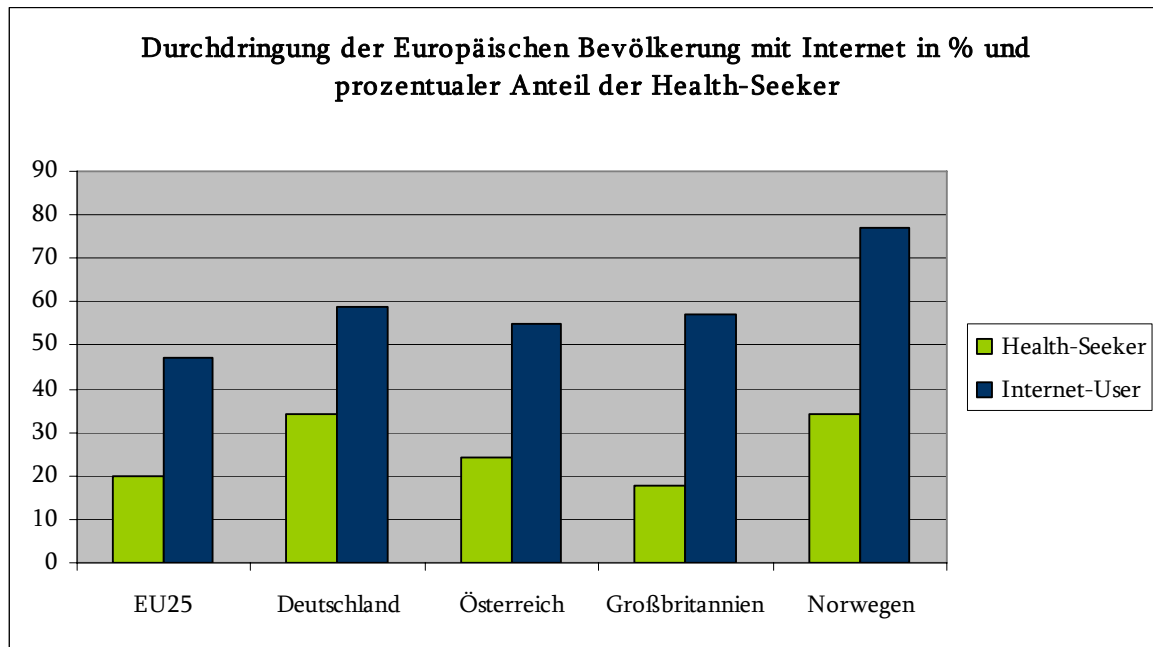


Abbildung 3: Internet-Durchdringung in Europa (Eurostat)

Im Europa der EU 25 liegt die Anzahl der Einzelpersonen, die Internet benutzen bei 47%. Die Verteilung in den einzelnen Ländern ist jedoch unterschiedlich ausgeprägt. Sie liegt in Deutschland bei etwa 60%. Auch in allen anderen Ländern, die aufgrund der Literaturrecherche berücksichtigt worden sind, liegen die Werte über dem EU25-Durchschnitt (Abbildung 3).

In Europa (EU der 25) benutzten laut Eurostat⁴² fast 90% aller Internet-User das Internet auch regelmäßig. Ungefähr 40 % aller Benutzer die älter als 24 Jahre sind, benutzen das Internet auch, um gesundheitsrelevante Informationen, und zwar für sich selbst oder für andere Personen, zu beschaffen⁴³. Bei der Generation der 16-24-Jährigen geht dieser Anteil auf 24 % zurück. Abbildung 4 zeigt die Verhältnisse im Detail.

⁴¹Vgl. Fox, Susanna *Online Health Search 2006* (2006)

<http://www.pewinternet.org/pdfs/PIP_Online_Health_2006.pdf> am 03.03.2007

⁴² Vgl. Datenauszug des Europäischen Amtes für Statistik, Eurostat im Anhang, S. 85.

⁴³ Vgl. Datenauszug des Europäischen Amtes für Statistik, Eurostat im Anhang, S. 85.

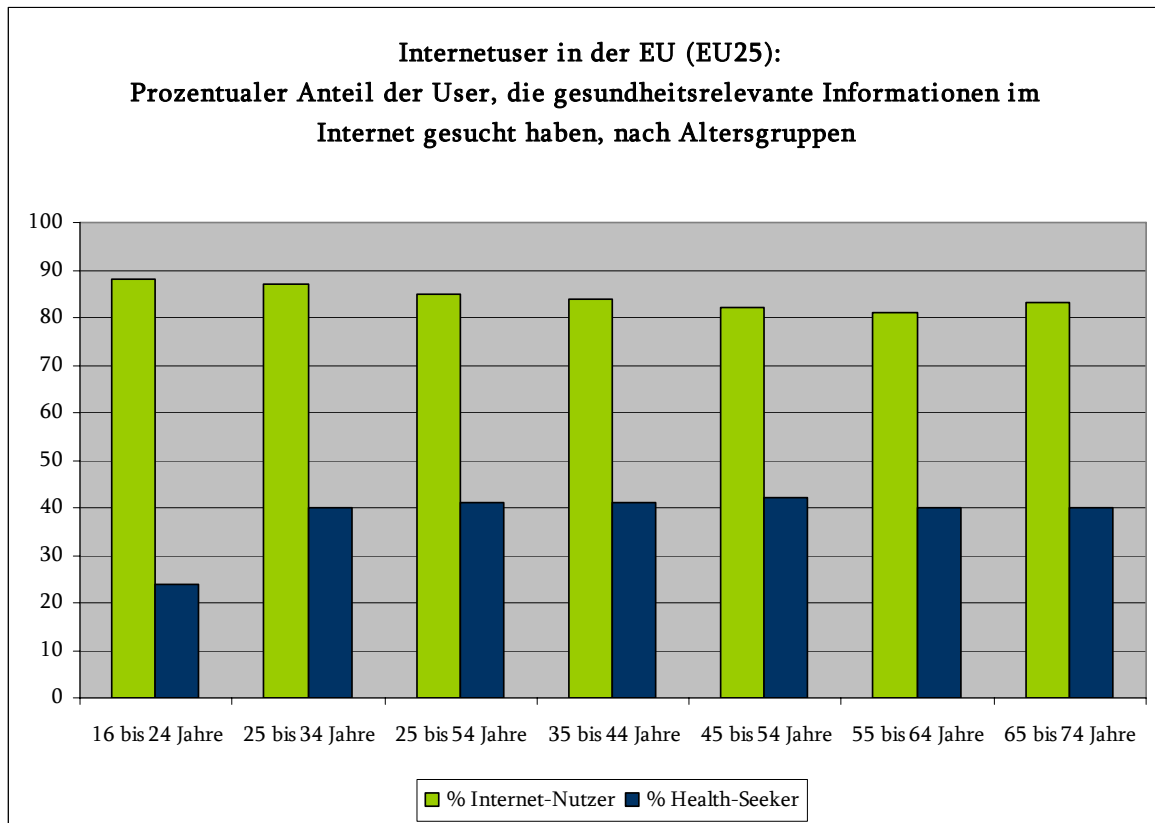


Abbildung 4: Anteil der Health-Seeker in Europa (Eurostat, 2007)

In Österreich haben im 4. Quartal 2006 67% aller Bürger Zugang zu Internet. 61% davon greifen auf aktuelle Informationen und auf Nachrichten zu, 50% auf Nachschlagewerke und wissenschaftliche Artikel. Während die 14-19-Jährigen zu 90% das Internet nutzen, sinkt dieser Anteil relativ gleichmäßig bis auf 57% in der Gruppe der 50-59-Jährigen, und fällt in der Gruppe der 60+ Jährigen stark auf 21%.⁴⁴

Die Daten von Eurostat zeigen für Österreich ein ähnliches Bild, der Anteil an Health-Seekern ist sehr ähnlich dem Anteil für die Staaten der EU 25 (Abbildung 5).

In einer norwegischen Studie über den Gebrauch des Internets für gesundheitsbezogene Suche nach Information wurden 1007 Personen über 15 Jahre telefonisch über ihr Verhalten befragt. 80% (804) Personen gaben an, das Internet regelmäßig zu verwenden, 58% (586 Personen) dieser Gruppe verwendeten das Internet auch zur Beschaffung von Gesundheitsinformationen.⁴⁵

⁴⁴ o.V., *Internet Monitor Austria - Kommunikation und IT in Österreich* (2006)

<http://www.integral.co.at/dlimages/AIM-C_4Quartal_2006.pdf> am 03.03.2007.

⁴⁵ Vgl. Andreassen, H. K. et al, 'Health-related use of the Internet in the Norwegian population' (2006) 126(22) *Tidsskr Nor Laegeforen* 2950-2952, S. 2951.

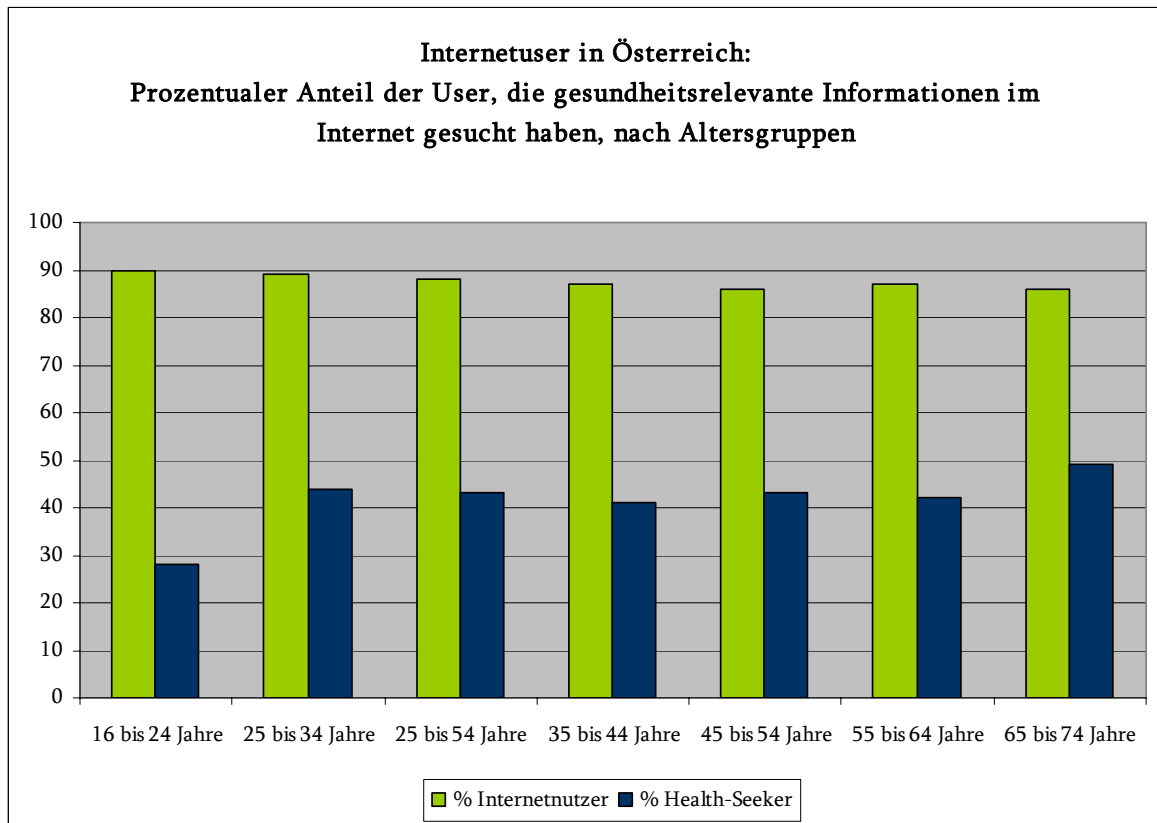


Abbildung 5: Anteil der Health-Seeker in Österreich (Eurostat, 2007)

3.1.3 Benutzung von Suchmaschinen

In den USA startet die Suche nach Gesundheitsinformation im Web an einer großen Suchmaschine. Zwei Drittel aller Suchenden beginnen mit einfachen Abfragebegriffen bei Suchdiensten wie Google oder Yahoo und besuchen einige wenige Websites mit Gesundheitsinformationen. 27% beginnen die Suche bei einer auf Gesundheitsinformationen spezialisierten Website⁴⁶.

Die wichtigsten Themen bei der Suche sind Informationen über eine spezifische Krankheit oder ein spezifisches medizinisches Problem (64% aller Nutzer). Verbreitet ist auch die Nachfrage nach einer bestimmten Behandlungsmethode (51%) und nach Informationen über Ernährungsfragen (49%).⁴⁷

Für Europa gibt es keine vergleichbar spezifischen Daten, jedoch wurden in kleineren Studien Personengruppen unabhängig voneinander über ihr Nutzungsverhalten befragt.

⁴⁶Vgl. Fox, Susanna *Online Health Search 2006* (2006)

<http://www.pewinternet.org/pdfs/PIP_Online_Health_2006.pdf> am 03.03.2007, S. 5.

⁴⁷ Vgl. Fox, Susanna *Online Health Search 2006* (2006)

<http://www.pewinternet.org/pdfs/PIP_Online_Health_2006.pdf> am 03.03.2007, S. 4.

Sim et al stellten etwa in einer 2007 veröffentlichten Studie folgendes Verhalten fest: Angehörige von kindlichen Patienten gaben zu 53% (N=271) an, das Internet für die Suche nach Gesundheitsinformationen für ihre Kinder zu verwenden. 75% davon verwendeten die Suchmaschine Google, um Informationen über den Gesundheitszustand ihrer Kinder zu beschaffen.⁴⁸

3.1.4 Zufriedenheit der Health-Seeker mit der Gesundheitsinformation

Die Zufriedenheit mit den dargebotenen Gesundheitsinformationen stufen Information suchende Laien immer höher ein als medizinisches Fachpersonal. Dies wurde von mehreren Autoren unabhängig festgestellt:

In einer aktuellen Untersuchung sagen etwa die Autoren einer Studie des Nottingham City Hospital: *„Patients do not have the knowledge to discern between good and poor quality Internet sites and tend to give higher ratings to the quality of health information obtained on the Internet than healthcare professionals.“*⁴⁹

Unterschiede treten bei der Analyse der Aufmerksamkeit der Health-Seeker auf, wenn sie darüber berichten, ob sie die Kennzeichnungen einer Webinformation mit Qualitätskriterien wie Quellennennung und Aktualisierungsdatum überprüfen.

In den USA hat die Aufmerksamkeit hinsichtlich des Überprüfens der Qualitätsmerkmale der Gesundheitsinformationen stark nachgelassen. 2001 lag der Anteil der Health-Seeker, die immer oder meistens die Herkunft und die Aktualität der Information überprüften bei 50%. 2006 sagen annähernd drei Viertel aller Health-Seeker, sie überprüfen solche Informationen selten bis gar nicht⁵⁰.

3.2 Begriffserklärungen

Im folgenden Kapitel werden die wesentlichen Begriffe, die mit dem Thema der Masterthese in Zusammenhang stehen, im Detail erläutert. Neben technischen Erklärungen sind hier auch soziale Phänomene berücksichtigt, die das Internet der letzten Jahre prägen.

⁴⁸ Vgl. Sim, Nurain Z. et al, 'Information on the World Wide Web—how useful is it for parents?' (2007) 42 *Journal of Pediatric Surgery* 305-312.

⁴⁹ Ayantunde, A. A. / N. T. Welch / S. L. Parsons, 'A survey of patient satisfaction and use of the Internet for health information' (2007) 61(3) *Int J Clin Pract* 458-462, S. 461.

⁵⁰ Vgl. Fox, Susanna *Online Health Search 2006* (2006)

<http://www.pewinternet.org/pdfs/PIP_Online_Health_2006.pdf> am 03.03.2007, S. 11.

3.2.1 Das Web 2.0

Der Begriff *Web 2.0* ist ein mehrdeutiger Begriff ohne exakte Definition. Er wurde erstmals 2004 als Teil des Namens einer Konferenz von *O'Reilly Media* (früher *O'Reilly & Associates*) verwendet. Mittlerweile ist der Begriff aber weit verbreitet und wird in mehreren Zusammenhängen verwendet, eine genauere Betrachtung seiner Bedeutungen ist daher erforderlich.

Web 2.0 bezeichnet eine allgemein wahrgenommene Änderung auf das Nutzungsverhalten des Internet und daher auf verschiedene Phänomene, die mit den stetig steigenden Nutzerzahlen des Internets sowie mit der Zunahme der Breitbandanschlüsse zusammenhängen.

Der Begriff wird aber auch technologiebezogen verwendet und überschneidet sich hier mit dem Begriff *Semantic Web*, mit dem er oft gleichgesetzt wird. Da *Web 2.0* Systeme aber mit aus dem World Wide Web bekannten Technologien arbeiten, ist diese Auslegung problematisch, weil sie an den wesentlichen Kriterien, die das *Web 2.0* ausmachen, vorbeigeht.

Vom Verlag, auf dessen Konferenz der Begriff erstmals verwendet wurde, dem O'Reilly Verlag, gibt es zwei Definitionen:

John Musser:

*“Web 2.0 is a set of economic, social, and technology trends that collectively form the basis for the next generation of the Internet - a more mature, distinctive medium characterized by user participation, openness, and network effects.”*⁵¹

Tim O'Reilly:

*“Web 2.0 is the business revolution in the computer industry caused by the move to the internet as platform, and an attempt to understand the rules for success on that new platform. Chief among those rules is this: Build applications that harness network effects to get better the more people use them. (This is what I've elsewhere called “harnessing collective intelligence”).”*⁵²

⁵¹ Musser, John, *Web 2.0 Principles and Best Practices* (2006)

⁵² O'Reilly, Tim, *Web 2.0 Compact Definition: Trying Again* (2006) O'Reilly Radar

<http://radar.oreilly.com/archives/2006/12/web_20_compact.html> am 02.03.2007 .

Das Web 2.0 ist nach diesen Definitionen stark von der Anzahl der Benutzer und von Netzwerkeffekten, wie sie Robert Metcalfe ursprünglich für die 1973 eingeführte Netzwerktechnologie *Ethernet* postulierte⁵³, abhängig.

Technologisch gesehen baut das *Web 2.0* auf den bewährten Fundamenten des World Wide Webs auf, ist also aus dieser Sicht betrachtet nicht innovativ.

Allerdings nimmt die Sicht, das World Wide Web als Plattform zu begreifen im Design von Web-Anwendungen einen immer breiteren Raum ein. Im Gegensatz dazu steht die herkömmliche Betrachtung, das Internet als Anhängsel eines Personal Computers oder eines lokalen Netzwerks zu sehen⁵⁴.

Die ursprünglich von Kritikern eingebrachte Sicht, der Begriff sei nichts anderes als ein Marketing-Schlagwort, ist heute nicht mehr vertretbar. Zu sehr werden damit konkrete Auswirkungen des durch die Masse der Nutzer veränderten Internet beschrieben.

Web 2.0 ist zu einem gebräuchlichen Begriff geworden, der den erweiterten Nutzen des Internet als Plattform heute treffend bezeichnet. Es existieren auch konkret Systeme⁵⁵, die sich solche Phänomene zu Nutzen machen.

3.2.1.1 Internet Social Networking

Social Networking ist die Bezeichnung für die Teilnahme großer sozialer Gruppen an kollaborativen Online-Systemen. Blogs, Wikis und Online-Communities sind Ausdruck für diese Entwicklung.

3.2.1.2 User Generated Content (UGC)

Eine wesentliche Erscheinungsform des Web 2.0 ist der durch Nutzer generierte Inhalt, der User Generated Content (UGC). Dabei werden von einer Vielzahl von Nutzern Inhalte in Form von Texten, Bildern, Videos oder anderen Informationen für die gemeinsame Nutzung in der Regel kostenfrei zur Verfügung gestellt. UGC trägt im Wesentlichen dazu bei, die Grenze zwischen Konsumenten und Produzenten aufzuweichen.

Ein Beispiel solcher kollaborativen Aktivität ist die Website <http://del.icio.us>, eine Online-Bookmarkverwaltung, die auch als *Social Bookmarking* bezeichnet wird.

⁵³ Auch bezeichnet als *Metcalfe's law*.

⁵⁴ Vgl. Antoniou, G. / Frank van Harmelen, *A semantic Web primer*, Cooperative information systems (2004), S. 3.

⁵⁵ Z. B. predictwallstreet.com, google.com.

3.2.1.3 Harnessing Collective Intelligence (HCI)

Eng im Zusammenhang mit UCG steht der sehr junge Begriff HCI, dessen Bedeutung aber wesentlich weiter gefasst wird. HCI leitet sich von CI, kollektive Intelligenz ab und bezeichnet die Nutzbarmachung der kollektiven Intelligenz aus Web 2.0-Systemen.

Eine Gruppe von Menschen, die in der Gesamtheit der Gruppe bessere Entscheidungen trifft, als das einzelne Mitglied der Gruppe, wird als kollektiv intelligent betrachtet. Nach diesem Prinzip arbeitet beispielsweise der Aktien-Dienstleister predictwallstreet.com, um Börsenkurse vorauszusagen.⁵⁶

3.2.2 Wikipedia

Wikipedia ist eine Online-Enzyklopädie, die auf die klassische redaktionelle Überwachung durch Editoren und Herausgeber verzichtet und stattdessen auf die kollektive Gestaltungskraft der Massen setzt.

Das Bemerkenswerte dabei ist nicht die zur Verwendung kommende Technologie an sich. Das System ist sehr einfach und wurde bereits vor der Entstehung der Wikipedia im Jahr 1995 von Ward Cunningham (*wiki wiki*⁵⁷) entwickelt. Die Haupteigenschaft der Wiki-Software ist die extreme Reduzierung des Aufwandes bei der Erstellung eines Inhaltes in einem kollaborativen System⁵⁸.

Seither ist es möglich, mit nur einem Webbrowser als Hilfsmittel, Inhalte zu editieren oder zu erstellen und Informationen mit anderen zu verlinken.

Historisch gesehen ist die Wikipedia die erfolgreiche Weiterentwicklung des Enzyklopädieprojekts Nupedia, das auf klassischer Begutachtung durch ein Editorenteam (Peer-Review) beruhte.

Die Nupedia Online-Bibliothek ist mit Wiki-Technologie entwickelt worden. Während Nupedia 2002 eingestellt wurde, hat sich Wikipedia, das im gleichen Labor wie Nupedia versuchsweise entstanden ist, zu einem funktionierenden System und schließlich zu einem globalen Phänomen entwickelt⁵⁹. Als Voraussetzung für die rasante Verbreitung

⁵⁶ Vgl. Kaplan, Craig A., 'Collective Intelligence: a new approach to stock price forecasting' (Arbeit präsentiert bei den Proceedings of the 2001 IEEE Systems, Man, and Cybernetics Conference, Tucson, AZ, USA, 2001), S. 2893.

⁵⁷ *wiki wiki*, hawaiianischer Ausdruck für schnell, flott.

⁵⁸ Vgl. Stvilia, Besiki et al, 'Assessing Information Quality of a Community-Based Encyclopedia' (Arbeit präsentiert bei den Proceedings of the 2005 International Conference on Information Quality (MIT IQ Conference), Cambridge, MA, USA, 2005).

⁵⁹ Voss, J. / P. Danowski, 'Bibliothek, Information und Dokumentation in der Wikipedia' (2004) 55(8) *Information, Wissenschaft & Praxis* 457-462, S. 458.

der Wikipedia gelten einerseits die einfache Bedienung und andererseits auch die frei verfügbare Open-Source-Software.

Die 2003 gegründete Wikipedia Foundation betreibt mittlerweile Enzyklopädien in 149 Sprachen, Wikipedia gehört derzeit zu den 20 bekanntesten Webseiten der Welt⁶⁰.

Die als Non-Profit-Organisation agierende Wikipedia Foundation hat aufgrund des rasanten Wachstums der Enzyklopädie die Aufgabe, Ressourcen für die Aufrechterhaltung des Betriebs zur Verfügung zu stellen.

Die deutsche Wikipedia ist neben der englischsprachigen die zweitgrößte Enzyklopädie und verfügte am 03.05.2007 über 578.915 Artikel bzw. eine Gesamtanzahl von 1.560.589 Seiten in der Datenbank. Seit 20.2.2002, dem Start der statistischen Aufzeichnungen von Wikipedia in Deutschland, wurden 32.628.704-mal Seiten bearbeitet, das entspricht durchschnittlich 20,91 Bearbeitungen pro Seite.⁶¹ Im Vergleich dazu liegen die durchschnittlichen Bearbeitungen der Krankheitsartikel der Primärprobe und Sekundärprobe der Feldanalyse bei 370 bzw. 176,93 pro Seite.

Im Mittelpunkt von Wikipedia steht die Mitarbeit Tausender Freiwilliger. Jeder Teilnehmer der Informationsgesellschaft kann in Wikipedia als Autor teilnehmen. Die Syntax zum Verfassen von Artikeln in Wikipedia ist sehr einfach und setzt keine HTML-Kenntnisse voraus⁶². Formatierungen, das Verlinken von Inhalten innerhalb und außerhalb der Wikipedia sind ebenso einfach und effizient machbar.

Wikipedia verzichtet auf ein übergeordnetes System der Inhaltskontrolle durch Editoren. Eine große Streuung in der Qualität der Inhalte wäre zu erwarten. Dennoch belegen verschiedene Studien, dass die Qualität in vieler Hinsicht bemerkenswert hoch ist⁶³. Bekannt ist auch, dass die Benutzung der Wikipedia vor allem unter Informatikern weit verbreitet ist.

⁶⁰Vgl. o.V., *Alexa Top 500 Web Sites* (2007) Alexa Internet, Inc.

<http://www.alexa.com/data/details/traffic_details?url=wikipedia.org> am 10.05.2007 .

⁶¹ Vgl. o.V., *Seitenstatistik* (2007) <<http://de.wikipedia.org/wiki/Spezial:Statistik>> am 03.05.2007 .

⁶² Vgl. Voss, J. / P. Danowski, 'Bibliothek, Information und Dokumentation in der Wikipedia' (2004) 55(8) *Information, Wissenschaft & Praxis* 457-462, S. 457.

⁶³ Vgl. Stvilia, Besiki et al, 'Assessing Information Quality of a Community-Based Encyclopedia' (Arbeit präsentiert bei den Proceedings of the 2005 International Conference on Information Quality (MIT IQ Conference), Cambridge, MA, USA, 2005), S. 2.

3.2.2.1 Das Revisionskontrollsystem der Wikipedia

Eine der wesentlichen Funktionen, die die Qualität der Information sichern hilft, ist das Revisionskontrollsystem. Pro Lemma werden in Wikipedia alle je durchgeführten Änderungen in einem Versionsregister gespeichert. Vorherige Versionen können so verglichen und bei Bedarf wieder zurückgesetzt werden. Dadurch steht den Autoren eine äußerst effektive Kontrolle zur Überwachung ihrer Artikel zur Verfügung.

3.2.2.2 Evolutive Artikelentwicklung (Selbstheilungskraft)

Die Entwicklung von den Anfängen eines Artikels zu einem reifen lexikalischen Werk wird auch als evolutive Artikelentwicklung im Sinne Darwins oder, im Falle von Schäden durch Vandalismus, als Selbstheilungskraft bezeichnet⁶⁴.

Ein Artikel entwickelt oder repariert sich durch die Vielzahl der Autoren mehr oder weniger schnell. Dieses Phänomen lässt sich durch einen Selbstversuch jederzeit nachvollziehen.

Eines der wenigen Prinzipien, die für Autoren der Wikipedia gelten, ist die Einhaltung des lexikalischen Stils, der zur ausgewogenen Informationsgestaltung verpflichtet.

3.2.2.3 Vandalismus

Als Vandalismus bei Wikipedia bezeichnet man die grund- und ziellose Löschung oder Veränderung von Inhalten durch Personen, die Freude am Verursachen von Schäden haben.

Diesem digitalen Phänomen wurde lange viel Bedeutung zugemessen, die große Zahl an konstruktiven Teilnehmern bewirkt aber, dass solche Zerstörungen sehr schnell aufgefunden und rückgängig gemacht werden können⁶⁵. Die Anzahl an Vandalen ist augenscheinlich nicht so hoch wie angenommen. Ein nachhaltiger Schaden durch den digitalen Vandalismus ist bis jetzt nicht zu beobachten.

3.2.2.4 Fälschung

Unter Fälschung wird in der Wikipedia vor allem das Beschönigen von Lebensläufen von im Mittelpunkt der Gesellschaft stehenden Menschen (vor allem von politischen Lebensläufen), sowie die Selbstdarstellung bzw. Selbstüberhöhung angesehen. Auch

⁶⁴ Vgl. Voss, J. / P. Danowski, 'Bibliothek, Information und Dokumentation in der Wikipedia' (2004) 55(8) *Information, Wissenschaft & Praxis* 457-462, S. 460.

⁶⁵ Vgl. Voss, J. / P. Danowski, 'Bibliothek, Information und Dokumentation in der Wikipedia' (2004) 55(8) *Information, Wissenschaft & Praxis* 457-462, S. 459.

Artikel, mit geschichtlichem Inhalt, wie etwa über jüngst vergangene Geschichte, können von politisch motivierten Fälschungen bzw. Verzerrungen betroffen sein.

3.2.3 Das Semantic Web

Semantic Web bezeichnet die Erweiterung des World Wide Web um maschinenlesbare Daten. Computer sollen in die Lage versetzt werden, die Bedeutung (Semantik) der Information zu verstehen und diese zu verarbeiten. Der Begriff geht ebenso wie die weltweit bekannten Technologiebegriffe URL, http und HTML auf Tim Berners-Lee zurück. Berners-Lee prägte den Begriff World Wide Web und ist ebenso Namensgeber als auch Vordenker des Semantic Web.

Die Mission der Semantic Web Initiative, entwickelt und kontrolliert im W3C, ist sehr breit gefasst. Wie das World Wide Web ein universelles Medium der Kommunikation ist, soll das Semantic Web das Medium des globalen Datenaustausches sein.

Das Semantic Web ist dabei nicht als neue Informationsplattform zu verstehen, die parallel zum aktuellen World Wide Web existiert. Es entwickelt sich vielmehr aus dem bestehenden Web heraus⁶⁶.

Ziel ist die reibungslose Zusammenführung von Daten aus Informationssystemen von Personen, Unternehmen und Organisationen und die weltweite, gemeinsame Nutzung von kommerziellen, wissenschaftlichen und kulturellen Daten. Die Informationen des aktuellen Webs, die eher schwach strukturiert sind, müssen in eine besser strukturierte Form übergeführt werden.

Erste sichtbare Applikationen des Semantic Web sind intelligente Agenten, die in der Lage sind, Information zu finden, zu sammeln und abzuleiten, um dem Benutzer exaktere, zeitgerechtere und sinnvollere Antworten zu liefern.

Das volle Potential kann das World Wide Web nur dann ausschöpfen, wenn Daten von Maschinen ebenso genutzt und verarbeitet werden können wie von Menschen.

Für das Web dieser Dimension ist es von eminenter Wichtigkeit, dass zukünftige Programme Daten aus unterschiedlichen Quellen nutzen und verarbeiten können, auch wenn diese Programme völlig unabhängig voneinander entwickelt worden sind. Wie das aktuelle Web baut auch das Semantic Web auf dem prinzipiellen Design der Dezentralisierung auf.

⁶⁶ Vgl. Antoniou, G. / Frank van Harmelen, *A semantic Web primer*, Cooperative information systems (2004), S.3.

Die wesentlichen Technologien des Semantic Web lassen sich in einem Schichtmodell, den Semantic Web Layer, darstellen. Abbildung 6 zeigt den Semantic Web Layer des W3C⁶⁷.

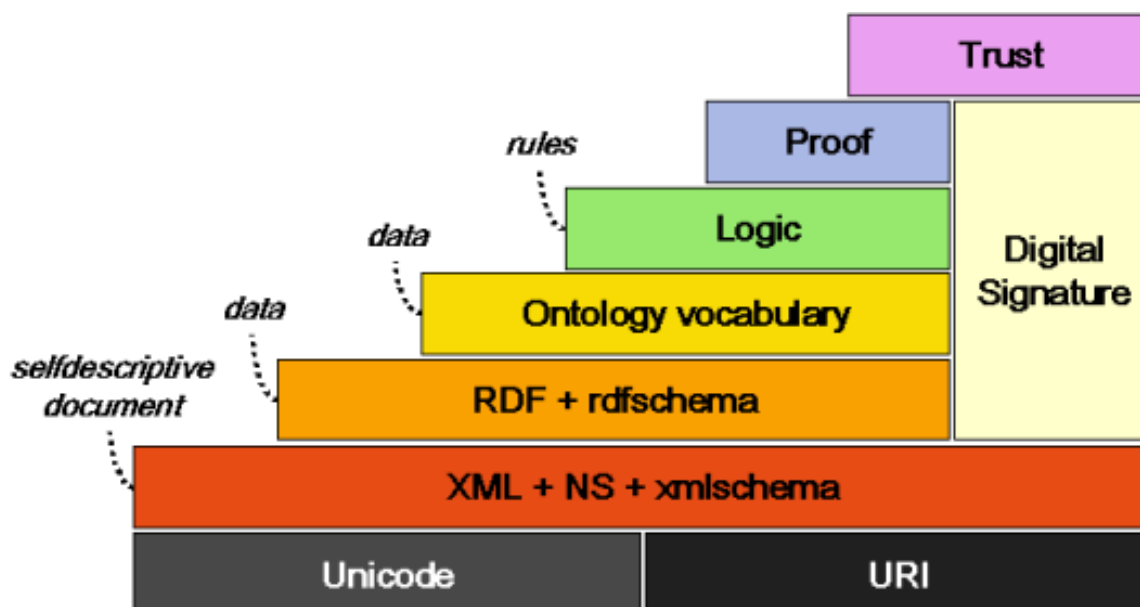


Abbildung 6: Semantic Web Layer (nach Koivunen / Miller 2001)

Die aktuellen Komponenten sind das RDF⁶⁸ Kernmodell, die RDF Schema Language und die Web Ontology Language (OWL). Da diese Modelle auch für die Entwicklung von Tools der Qualitätsinitiativen für Gesundheitsinformationen von grundlegender Bedeutung sind, wird in den nächsten vier Abschnitten näher darauf eingegangen.

3.2.3.1 Metadaten

Als Metadaten bezeichnet man präzise, übergeordnete Informationen über alle Typen von Daten. Im World Wide Web wird der Begriff im Speziellen in folgendem Zusammenhang verwendet:

(1) Beschreibende Information zur Verbesserung der Kontrolle und der Auffindbarkeit von Information, (2) Beziehung von Inhalt, Struktur und logischen Informationen untereinander, (3) Verbesserung der Navigation im Web.⁶⁹

⁶⁷ Vgl. Koivunen, Marja-Riitta / Eric Miller, 'W3C Semantic Web Activity' (Arbeit präsentiert bei den Semantic Web Kick-off Seminar in Finland, Helsinki, 2001).

⁶⁸ RDF=Resource Description Framework.

⁶⁹ Vgl. Berners-Lee, Tim / Mark Fischetti, *Weaving the Web : The Past, Present and Future of the World Wide Web by its inventor* (1999), S. 195.

3.2.3.2 Repräsentation der Information (XML)

Information kann definiert werden als die Repräsentation von Wissen (Daten, Text, Numerische Daten, Bilddaten, audiovisuelle Daten) in jeglichen Medium.

Im World Wide Web liegt der Großteil der Informationen in unstrukturierter Form, heute zumeist noch in Form einfacher HTML-Dateien vor.

Eine deutliche Verbesserung der Darstellungsmöglichkeiten wurde 1996 mit der Einführung von CSS (Cascaded Style Sheets Level 1) erreicht. Andere Skriptsprachen, wie etwa PHP (Hypertext Preprocessor), erhöhten den Komfort der Erstellung von Webinhalten abermals, änderten aber wenige daran, dass weithin größtenteils unstrukturierte Daten das Web ausmachten.

Der Inhalt ist somit mehr nach den Bedürfnissen des menschlichen Benutzers gestaltet, als nach der Möglichkeit, Daten maschinell bzw. programmorientiert zu verarbeiten.

Dies änderte sich erst mit der Entwicklung von Metasprachen, die der Information selbst eine zusätzliche Ebene der Bedeutung hinzufügten.

Eine der bekanntesten Metasprachen zur Beschreibung von strukturierten Daten ist XML. Dokumente, die in XML abgefasst sind, können Strukturinformationen und Daten enthalten.

XML wurde vom W3C als einfach zu implementierendes Subset zu SGML⁷⁰ entwickelt und ist rasch zu einem Quasi-Standard für Austausch von strukturierten Daten aufgestiegen.

XML kann für unterschiedliche Zwecke eingesetzt werden. Ein Zweck etwa ist das Markup von Web-Dokumenten im Zusammenhang mit Style Sheets, um Information auf unterschiedlichen Plattformen (z. B. PC, Handhelds, Mobiltelefone) optimal darstellen zu können. Ein anderer Zweck ist die Definition eines Datenaustauschformats zwischen verschiedenen Applikationen.

3.2.3.3 Repräsentation des Wissens (RDF)

Das Resource Description Framework ist ursprünglich als Modell zur Beschreibung von Metasprachen spezifiziert worden, wird aber jetzt als generelle Methode zur Beschreibung von Information und Informationsflüssen für eine Vielzahl von Beschreibungsformaten verwendet⁷¹.

In einer einfachen Ausführung ist es ein einfaches relationales Modell und basiert auf XML. Es dient als Modell für die Darstellung von Daten, Objekten und

⁷⁰ SGML = Standard Generalized Markup Language: Metasprache zur Auszeichnung verschiedener Markup-sprachen, ISO Standard: *ISO 8879:1986*.

⁷¹ Vgl. o.V., *Resource Description Framework* (2007)

<http://en.wikipedia.org/wiki/Resource_Description_Framework> am 22.04.2007 .

Objektbeziehungen und erlaubt die Verteilung und die gemeinsame Verwendung von strukturierten und semistrukturierten Daten über Applikationsgrenzen hinaus⁷². Während XML nur Dokumente beschreiben kann, können RDF-Daten verschiedenartigste Objekte beschreiben.

Eine Alternative zum RDF wurde 1999 in einer ISO/IEC 13250:2003 Norm standardisiert. In dieser Norm wird die Wissensrepräsentation mit einer über den Objekten stehenden Sicht mit einer Gewichtung auf Auffindbarkeit, den so genannten Topic Maps, dargestellt. Die überwiegende Anzahl der aktuellen Entwicklungen im Semantic Web finden aber im weniger strikt regulierten RDF statt.

3.2.3.4 Spezifizierung des Wissens (RDF Schema, OWL)

Die Repräsentation von Information sowie die Einbettung der Wissensbegriffe in ein Format, das von Maschinen verstanden und verarbeitet werden kann, setzt die Verständigung auf ein gemeinsames Vokabular, eine gemeinsame Systematik und eine Ontologie voraus⁷³.

Das RDF-Schema erweitert das RDF um ein Modellkonzept mit Semantik und Regeln für Ableitungen und Rückschlüsse. Berners-Lee nennt zur Illustration dafür einen analogen Vergleich:

„If HTML and the Web made all the online documents look like one huge book, RDF, schema, and inference languages will make all the data in the world look like one huge database.“⁷⁴

3.2.3.5 Ontologien

Die Computerwissenschaft hat den ursprünglich aus der Philosophie stammenden Begriff übernommen.

Die Ontologie in der Philosophie ist ein Fachgebiet, das sich mit der Lehre des Wesens der Existenz, der existierenden Dinge und ihrer Beschreibung und Kategorisierung beschäftigt, die aus dem griechischen Altertum stammende *ὄντολογία* ist ein Teilgebiet der Metaphysik.

⁷² Vgl. Berners-Lee, Timothy et al, *Frequently Asked Questions about RDF* (2004)
<<http://www.w3.org/RDF/FAQ>> am 20.04.2007 .

⁷³ Vgl. Wong, Chris, *XML, RDF and the Semantic Web: application to knowledge management* University of Technology, Sydney, 2003).

⁷⁴ Berners-Lee, Tim / Mark Fischetti, *Weaving the Web : The Past, Present and Future of the World Wide Web by its inventor* (1999), S. 201.

In der Computerwissenschaft ist der Begriff jedoch mit einer spezifischeren, technischen Bedeutung assoziiert.

Für das Funktionieren des semantischen Suchens in Daten oder bei der Einholung von Information aus dem Web müssen gemeinsame Vereinbarungen über die Bedeutung und die Beziehung der Begriffe vorhanden sein. Ontologien stellen solche Vereinbarungen dar. Sie beschreiben die formalen Methoden für die allgemeine Darstellung und Bedeutungen eines Schemas oder einer Wissensdomain.

Ein Beispiel für eine umfangreiche Ontologie ist das UMLS, das Unified Medical Language System. Das 1989 von der National Library of Medicine initiierte Projekt hat zu Ziel, die Interoperabilität heterogener medizinischer Quellsysteme in der Medizindomain zu erhöhen.

Ein Beispiel für eine minimale Variante einer Ontologie mit nur 15 grundlegenden Metadatenelementen wurde von der Dublin Core Metadata Initiative⁷⁵ - ursprünglich für die Bibliothekswissenschaften - entwickelt. Beide, UMLS und Dublin Core, spielen für die Entwicklung von Tools für die Qualitätsbewertung von Gesundheitsinformationen eine Rolle.⁷⁶

3.2.3.6 Exkurs: Ein Tag mit Michaels Semantic Web Agenten

Wenn einst Ontologien in ausreichendem Maß im Web vorhanden sind, werden solche zukünftigen Szenarien, wie unten beschrieben, automatisch von Software-Agenten ausgeführt:

Michael hatte eben einen kleinen Verkehrsunfall, der ihm Nackenschmerzen bereitete. Sein erstversorgender Arzt stellte eine Überdehnung fest und empfahl einige Massagen. Michael fragte seinen Semantic Web Agenten, um einige Möglichkeiten der Behandlung auszuarbeiten.

Der Agent erhielt Details über die vorgeschlagene Therapie vom Agenten des erstversorgenden Arztes und fragte die Liste der Physiotherapeuten ab, die Michaels Krankenversicherung akzeptieren. Der Agent schränkte die Suche auf einen Radius von 10 km von Michaels Arbeitsplatz und Wohnung ein. Dann versuchte er, mögliche Behandlungstermine mit Michaels Kalender abzuklären. Nach einigen Minuten lieferte der Agent zwei Angebote von Physiotherapeuten zurück, die Michael aber nicht

⁷⁵ DCMI, Webseite: <http://dublincore.org/>.

⁷⁶ Vgl. Kap. 3.4, S. 47.

überzeugten. Beim ersten Arzt hätte er zwei Wochen Wartezeit und beim zweiten müsste er zur verkehrsreichsten Zeit mitten durch die Stadt fahren.

Michael fragte seinen Agenten noch einmal, diesmal mit einem strikteren Zeitplan. Nach einigen Minuten lieferte der Agent ein neues Angebot zurück. Ein bekannter Physiotherapeut ohne Kassenvertrag bot schon in zwei Tagen einen ersten Termin zur Massage an.

Würde er dieses Angebot annehmen, müsste Michael dafür allerdings einige seiner weniger wichtigen Termine verschieben, und der Agent bot an, diese Terminverschiebungen durchzuführen, würde er sich für diesen Arzt entscheiden.

Der Physiotherapeut hatte zwar keinen Vertrag für die Kostenübernahme von Michaels Versicherung, der Agent hatte aber bereits geprüft, dass Michaels Versicherung den Pauschalbetrag für die Behandlung übernehmen würde. Auch hat er bereits mit dem Agenten des Physiotherapeuten einen Preisnachlass verhandelt, sodass für Michael kaum Mehrkosten entstehen würden.

Michael war froh und akzeptierte das Angebot, war aber auch neugierig, wie denn der Agent zu seinen Entscheidungen gekommen ist, hatte er den Semantic Web Agenten doch erst vor wenigen Tagen installiert. Wie wurde die Reputation des Physiotherapeuten ermittelt? Wie erkannte der Agent, welche von Michaels Terminen zu den weniger wichtigen gehörten und wie wurde die Preisverhandlung mit dem Physiotherapeuten durchgeführt? Der Agent lieferte die entsprechenden Informationen prompt zurück an Michael.

Michael war zufrieden, sein Web Agent schickte sich an, sein hektisches Leben etwas einfacher zu gestalten. Er beauftragte den Agenten, alle notwendigen Maßnahmen zu setzen, um die Aufgabe zu finalisieren.⁷⁷

Dieses Szenario ist nicht Science-Fiction. Die Tools dafür sind vorhanden und es benötigt keine wissenschaftliche Forschung, um diese Ziele zu erreichen. Die Herausforderung liegt vielmehr in der Aneignung der Technologie durch Softwareentwickler und Benutzer.

⁷⁷ Vgl. Antoniou, G. / Frank van Harmelen, *A semantic Web primer*, Cooperative information systems (2004), S. 6.

3.3 Entwicklungstand der Gesundheits-Qualitätsinitiativen

Ahmad Risk⁷⁸ und Petra Wilson⁷⁹ haben die verschiedenen Qualitätsinitiativen im World Wide Web unabhängig voneinander anhand der Benutzung ihrer Technologien in fünf Gruppen eingeteilt, die in der Folge kurz dargestellt werden.

3.3.1 Codes of Conduct

Die gemeinsame Grundlage der *Codes of Conduct* - Verhaltenskodizes - sind die ethischen Grundlagen des menschlichen Handelns. Vor allem in der Frühzeit des kommerzialisierten Internets ist das Web mit Gesundheitsinformationen mit oft zweifelhafter Herkunft und Absicht überschwemmt worden.

Oft ließen sich Inhalte mit mangelhafter Qualität schwer von solchen mit guter Absicht und guter Qualität unterscheiden. Schon früh erkannten Mediziner und medizinische Organisationen Handlungsbedarf und begannen Maßnahmen zu setzen⁸⁰, gleichzeitig entwickelten sich die Verhaltenskodizes. Die Glaubwürdigkeit vieler Webseiten, im Speziellen auch vieler mit Inhalten zur Gesundheit wurde hinterfragt und es wurden auch bereits Zweifel an so manchen Qualitätsinitiativen selbst erhoben⁸¹.

Die wichtigsten Vertreter dieser Kategorie sind die bereits genannte HON (gegründet 1996), die Richtlinien der AMA⁸² (2000), und der eHealth Code of Ethics der Internet Healthcare Coalition⁸³ (2000). Ebenso wurden mit der Initiative eEurope 2002 der Europäischen Kommission Richtlinien für Anbieter und Konsumenten von Gesundheitsinformationen empfohlen, die auch als Verhaltenskodex bezeichnet werden können. Staatliche Regulierung bzw. der Aufbau von qualitätsgeprüften Gesundheitsportalen wurde vor allem seitens der EU angedacht, es blieb aber bisher bei Empfehlungen, die Mitgliedsstaaten sind nicht verpflichtet, diese Empfehlungen zu übernehmen.

⁷⁸ Vgl. Risk, Ahmad / Joan Dzenowagis, 'Review of internet health information quality initiatives' (2001) 3(4) *J Med Internet Res* E28.

⁷⁹ Vgl. Wilson, Petra, 'How to find the good and avoid the bad or ugly: a short guide to tools for rating quality of health information on the internet' (2002) 324 *Bmj* 598-602.

⁸⁰ Vgl. Jadad, Alejandro R Gagliardi, Anna, 'Rating Health Information on the Internet - Navigating to Knowledge or to Babel?' (1998) 279 *Jama* 611 - 614, S. 611.

⁸¹ Vgl. McLellan, Faith, 'Like hunger, like thirst: patients, journals, and the internet' (1998) Volume 352, Supplement 2 *The Lancet* S39-S43. S. 39.

⁸² American Medical Association.

⁸³ <http://www.ihealthcoalition.org>.

3.3.2 Gütesiegelsysteme

Qualitätssiegelsysteme, Gütesiegelsysteme - im Englischen verbreitet als Trustmark-Systeme bezeichnet - dienen vor allem dem schnellen Nachweis, dass dem gekennzeichneten Inhalt des Informationsanbieters vertraut werden kann.

Der Informationsanbieter übernimmt bestimmte Regeln der Vergabeinstitution des Gütesiegels und erhält im Gegenzug die Berechtigung zum Führen des Gütesiegels. Aufgrund ihrer einfachen Benutzbarkeit sind solche Systeme die am meisten akzeptierten und daher ebenso weit verbreitet. Hierzu zählen vor allem die HON, die spanische Web Medica Acredidata⁸⁴ und das Qualitätssiegelsystem des deutschen AFGIS⁸⁵.

3.3.3 Anleitungstools für Nutzer

Anleitungstools bieten dem Nutzer Hilfestellung selbst zu beurteilen, ob die dargebotene Information glaubwürdig ist. In vielen Fällen leiten diese Tools einen Benutzer zu einer Website weiter, wo anhand eines Fragebogens oder anhand einer Kriterienliste Unterstützung zur Selbstbeurteilung der Qualität geboten wird. Auch hier ist die Grundlage ein Verhaltenskodex. Wichtige Vertreter dieser Kategorie sind das englische Discern⁸⁶ und das französische NetScoring, letztgenannte Initiative ist nicht mehr aktiv bzw. hat sie sich mit zwei anderen Initiativen zu MedCircle⁸⁷ zusammengeschlossen und weiterentwickelt.

3.3.4 Filtertechnologien

Diese Technologien arbeiten oft mit Metadaten und Gateway-Architekturen. Sie richten sich eher an medizinisches Fachpersonal und Forscher und sind weniger für Laien geeignet. In der einschlägigen Literatur wurden sie aber stets zu den Qualitätsinitiativen gezählt.

3.3.5 Zertifizierung durch Dritte - Third Party Rating

Diese Methode versucht Mängel von leicht zu manipulierenden Gütesiegelsystemen auszugleichen, indem die Gesundheitsinformationen entweder über Gateway-Technologien oder über Metadaten Systeme durch Dritte, also von Fachleuten, auf Glaubwürdigkeit geprüft werden.

⁸⁴ <http://wma.comb.es/eng/presentacio.htm>.

⁸⁵ Aktionsforum Gesundheitsinformationssystem, <http://www.afgis.de>.

⁸⁶ <http://www.discern.org.uk>, <http://www.discern.de>.

⁸⁷ Vgl. Kap. 3.4.3, S. 48.

Systeme dieser Kategorie sind aufgrund des großen organisatorischen und personellen Aufwands sehr teuer in der Erhaltung und haben den Experimentalstatus oft nicht verlassen.

Als Upstream-Filtering wird hier die Selektion durch Dritte, meist eine Prüfung durch medizinische Fachleute verstanden. Diese Technologien haben sich nicht bewährt, um sich den dynamisch verändernden Inhalten des Webs anzupassen. Zusätzlicher Aufwand entsteht auch für die Gruppe der Benutzer, weil neben der eigentlichen Ressource der Gesundheitsinformation auch noch die Metainformation auf der Website des prüfenden Dritten berücksichtigt werden muss.⁸⁸

Unter Downstream-Filtering wird die Unterstützung des Benutzers bei der Beschaffung von Gesundheitsinformation mit Software Agenten verstanden.⁸⁹

Eine wichtige amerikanische Initiative dieser Ausprägung ist URAC⁹⁰, eine unabhängige Organisation, die Akkreditierungen und Zertifizierung für gesundheitsbezogene Websites anbietet. URAC setzt sich aus Vertretern der wesentlichen amerikanischen Gesundheitsorganisationen und Gesundheitsbehörden zusammen.

Ein wichtiges europäisches, mit Metadaten operierendes, System ist das EU-finanzierte MedCIRCLE Projekt. Diese Initiative hat schon frühzeitig begonnen, Semantic Web Technologien zu entwickeln und wird daher an entsprechender Stelle⁹¹ behandelt.

3.3.6 Medizinspezifische Suchmaschinen

Ein seit ungefähr drei Jahren zu beobachtender Trend im Internet ist die Entwicklung von vertikalen Suchmaschinen, die im Gegensatz zu allgemeinen Suchmaschinen nur einen spezifischen Teil des Webangebots durchsuchen und wiedergeben. Die Technologie einer spezialisierten Suche ist zwar nicht neu, aber die Verbreitung solcher vertikalen Suchmaschinen hat in den letzten Jahren stark zugenommen.

Dieser Trend hat bereits Anbieter von Suchmaschinen auf den Plan gerufen, die eine gesundheitspezifische Suche anbieten. Ein im englischsprachigen Raum bekannter Vertreter ist Healthline.com⁹². Dieser Anbieter sucht in seinen eigenen Daten sowie in

⁸⁸ Vgl. Eysenbach, Gunther Diepgen, Thomas L, 'Towards quality management of medical information on the internet: evaluation, labelling, and filtering of information' (1998) 317 *Bmj* 1496-1502, S. 1496-1497.

⁸⁹ Vgl. Eysenbach, Gunther Diepgen, Thomas L, 'Towards quality management of medical information on the internet: evaluation, labelling, and filtering of information' (1998) 317 *Bmj* 1496-1502, S. 1498.

⁹⁰ Ursprünglich stand URAC für das Akronym *Utilization Review Accreditation Commission*, <http://www.urac.org/>.

⁹¹ Vgl. Kap. 3.4.1, S.47.

⁹² Beispiel einer vertikalen Suchmaschine für Health Information, www.healthline.com.

Quellen von ausgewählten Gesundheitsdiensten und liefert als Ergebnis einer Abfrage immer zuerst die eigenen Inhalte, dann die Inhalte der externen Anbieter zurück. Externe Anbieter, falls sie über ein Zertifikat oder ein Gütesiegel einer Qualitätsinitiative verfügen, werden besonders hervorgehoben.

Im Jahre 2006 wurde auf vom Suchdienst Google eine entsprechende Initiative angekündigt, bisher ist aber noch keine praxistaugliche Applikation verfügbar.

Andere medizinspezifische vertikale Suchdienste, teilweise noch im Beta-Stadium, sind beispielsweise kosmix.com, mammahealth.com und medstory.com.

3.4 Semantic Web Technologien bei Gesundheitsinformationen

Bereits vor dem Jahr 2000 gab es erste Versuche, durch Metadatenmodelle eine Strukturierung der Webinformationen zu erreichen. Ziel war, eine leichtere Auffindbarkeit von Informationen, eine verbesserte Interoperabilität zwischen einzelnen Informationssystemen und vor allem auch eine leichtere Kontrolle der Inhalte zu ermöglichen.

3.4.1 UMLS

Das Unified Medical Language System hat zum Ziel, den Zugang zu der unüberblickbaren Menge an biomedizinischer Information im Internet zu verbessern. Es besteht aus einer Sammlung einer Vielzahl von Vokabularen aus dem medizinischen und biomedizinischen Umfeld und einer Metadatenstruktur, die diese Vokabularen verbindet. Heute bezeichnet man UMLS treffend als die grundlegende Ontologie für die Domain der Biomedizin. Sie baut auf drei hauptsächlichen Anwendungen auf.

3.4.1.1 Metathesaurus

Der Metathesaurus® ist die Basis des UMLS. Er umfasst über eine Million biomedizinischen Konzepte und 5 Millionen Konzeptnamen aus über 100 kontrollierten Vokabularen, darunter zum Beispiel MeSH, SNOMED, LOINC.

Der Metathesaurus stellt auch den grundlegenden Zusammenhang zwischen den Beziehungen der einzelnen Konzepte dar.⁹³

⁹³ Vgl. o.V., *NLM Fact Sheet UMLS Metathesaurus*

<<http://www.nlm.nih.gov/pubs/factsheets/umlsmeta.html>> am 25.04.2007

3.4.1.2 Semantisches Netzwerk

Das Semantische Netzwerk bei UMLS bezeichnet die Schemata zur Wissensrepräsentation der Objekte (Konzepte) und die Beziehung zwischen diesen Objekten. Die graphische Darstellung trägt zum Verständnis der Beziehungen bei. Im Semantischen Netzwerk werden die Konzepte aus dem Metathesaurus kategorisiert und die Kategorien wiederum in Beziehung zueinander gesetzt.

3.4.1.3 Specialist Lexicon

Das Specialist Lexicon wurde entwickelt, um für das SPECIALIST Natural Language Processing (NLP) System lexikalische Information bereitzustellen. Die Grundidee ist ein englisches Lexikon mit vielen biomedizinischen Fachbegriffen. Die Einträge im Lexikon zeichnen für jedes Wort oder für jeden Begriff die syntaktische, morphologische und orthografische Information auf, die vom SPECIALIST NLP System gebraucht wird.⁹⁴

3.4.2 Dublin Core

Das Metadatenet der Dublin Core Metadata Initiative geht auf das Jahr 1995 zurück, als in Dublin, Ohio, der erste Kongress von Bibliothekswissenschaftlern sich um ein Format für die verbesserte Interoperabilität bemühte.

Dublin Core ist heute standardisiert und dient in erster Linie der Beschreibung von Informationsressourcen über Domaingrenzen hinaus. Wichtiges und häufiges Anwendungsgebiet ist die Klassifizierung und Beschreibung zahlreicher Datenformate im World Wide Web. Dublin Core wird im Allgemeinen in XML und RDF ausgedrückt. Aus dem ursprünglichen einfachen Dublin Core Metadata Element Set (DCMES), das aus 15 Elementen bestand, entwickeln sich unter der Kontrolle der Initiative weitere Elemente bzw. werden bestehende Elemente verfeinert⁹⁵.

3.4.3 MedCIRCLE

Die MedCIRCLE Initiative, früher MedCERTAIN⁹⁶, ist eine Vereinigung von europäischen Einrichtungen, die seit 1998 versucht, Tools für die Bewertung von Gesundheitswebsites durch Dritte zu entwickeln. MedCIRCLE selbst tritt nicht als Bewertungs-Initiative auf. Der Schwerpunkt der Arbeit liegt auf der Entwicklung von Bewertungssystemen, um Laien und Konsumenten sichere Informationsquellen zu liefern. Das Projekt wurde bis

⁹⁴ Vgl. o.V., *NLM Fact Sheet SPECIALIST Lexicon* <<http://www.nlm.nih.gov/pubs/factsheets/umlslex.html>> am 25.04.2007 .

⁹⁵ Vgl. o.V., *Dublin Core (2007)* <http://de.wikipedia.org/wiki/Dublin_Core> am 25.04.2007 .

⁹⁶ Abkürzung für *MedPics Certification and Rating of Trustworthy Health Information on the Net*.

zum Jahr 2004 von der EU im Rahmen des Aktionsplans für „Sicherheit im Internet“ gefördert.

Ein weiteres Ziel ist, vor allem seit Semantic Web Technologien eingesetzt werden, die Interoperabilität zwischen Anbietern, Verbrauchern und Bewertungseinrichtungen zu erhöhen.

Die Mitglieder des MedCIRCLE Projekts setzten sich 2002 aus der Universität Heidelberg (Department of Clinical Social Medicine), der COMB (Col legi Oficial de Metges de Bacrlona), CISMef (Universität von Rouen, Frankreich) und der AQUMED (Agentur für Qualität in der Medizin) zusammen⁹⁷. Ein führender Wissenschaftler in dem Projekt war Gunther Eysenbach.

Bereits 1997 setzte MedCERTAIN ein als MedPICS bezeichnetes Metadatenet ein, das auf der W3C-Empfehlung PICS beruhte. Beabsichtigt waren beschreibende Metadaten wie Urheberschaft, Qualifikation der Autoren, Finanzierungsquellen, Schlüsselwörter betreffend des Inhalts, die auf UMLS, aber auch anderen Beschreibungssprachen beruhten.

MedCERTAIN wählte dieses Metadatenet, weil es laut Ansicht der Autoren im Gegensatz zum einfachen, beschreibenden System der Dublin Core Metadatenets erlaubte, dass Metadaten von Dritten übernommen werden können. Schon im Jahr 2000 kündigte MedCERTAIN die Migration auf XML/RDF Technologie an, da auch der PICS Standard zu dieser Zeit in die aufkommenden Semantic Web Technologien übergeführt wurde⁹⁸.

3.4.3.1 HIDDEL Vokabular für Metadaten

HIDDEL⁹⁹ ist der Name des Metadaten-Vokabulars, das von MedCERTAIN begonnen und unter MedCIRCLE weiterentwickelt wurde. Es wurde für die Verwendung als Fachgateway zur Qualitätssicherung im Sinne des Third-Party-Ratings entwickelt. AQMED, COMB und CISMef fungieren als Gateways für Gesundheitsinformationen (consumer health information), die die Informationen mittels HIDDEL Vokabular auf Glaubwürdigkeit überprüfen und Anmerkungen darüber speichern.

⁹⁷ Vgl. Mayer, M. A. et al, 'MedCIRCLE: collaboration for Internet rating, certification, labelling and evaluation of health information on the World-Wide-Web' (2003) 95 *Stud Health Technol Inform* 667-672, S. 667.

⁹⁸ Vgl. Eysenbach, Gunther et al, 'MedCertain: Quality Management, Certification and Rating of Health Information on the Net' (Arbeit präsentiert bei den AMIA 2000 Annual Symposion, Santa Anita, A, 2000)

⁹⁹ Abkürzung für *Health Information Disclosure, Description and Evaluation Language*.

Die Anmerkungen werden in XML/RDF ausgedrückt und öffentlich zugänglich gemacht. Intelligente Semantic Web Agents sind dann in der Lage, Informationen über die Glaubwürdigkeit von gesundheitsbezogenen Websites einzuholen.

3.4.4 HIQuA

Ein anschauliches Beispiel für die Anwendung von semantischen Konzepten bei der Suche nach Gesundheitsinformationen im Web (Health Information Retrieval) wurde von der Gruppe um Qing T. Zeng entwickelt.

Der so genannte Health Information Query Assistant (HiQuA) hilft den Konsumenten, besser artikuliert Suchabfragen zu formulieren, indem er alternative oder ergänzende Suchanfragen empfiehlt.¹⁰⁰

Ausgehend davon, dass Konsumenten aufgrund begrenzter medizinischer Sachkenntnis oft nicht nur sehr kurze, sondern auch nicht ausreichend spezifizierte Begriffe zur Informationsbeschaffung verwenden, wurde ein 3-stufiges Webtool (Java, Apache Tomcat Webserver, MySQL Datenbank) entwickelt. Das Bedienfenster der Anwendung wird bei der Abfrage einer Suchmaschine ergänzend in das Fenster des Webbrowsers eingeblendet.

Die Hauptfunktion des Systems ist es, medizinische Konzepte, die eine semantische Beziehung zum ursprünglichen Abfragestring des Benutzers aufweisen, zu erkennen und diese dem Benutzer zu empfehlen.

Das Umsetzen des Abfragestrings in medizinische Konzepte ist die wesentliche Neuerung, die mit dieser Technologie eingeführt wird. Ein jeder Abfragestring kann in ein oder mehrere Konzepte umgesetzt werden.

Die Konzepte sind im UMLS¹⁰¹ definiert und sollten eine enge Beziehung zu den ursprünglichen Abfragekonzepten, also eine kurze semantische Distanz zueinander haben.

Für die analytische Schätzung der semantischen Distanz werden drei Ressourcen verwendet. Als Ressource für medizinische Vokabularien wird die UMLS Metathesaurus Relationship (MRREL) Tabelle verwendet.

Zur Komplettierung der medizinischen Vokabularien wird das gleichzeitige Auftreten von Daten in Konzepten der medizinischen Literatur herangezogen. Dafür wird die UMLS Metathesaurus Co-Occurrence (MRCOC) Tabelle verwendet.

¹⁰⁰ Vgl. Zeng, Q. T. et al, 'Assisting consumer health information retrieval with query recommendations' (2006) 13(1) *J Am Med Inform Assoc* 80-90, S. 80.

¹⁰¹ Vgl. Kap. 3.4.1, S.47.

Die dritte Ressource ist das Einbinden der Konzepte in die Abfragesitzungen der Konsumenten (consumer HIR sessions).

Die Log-Daten dafür stammen aus einer amerikanischen Gesundheitswebseite für Konsumenten, der MedlinePlus¹⁰², die von der NLM betrieben wird.¹⁰³ Abbildung 7 gibt das grundlegende Design des HIQuA wieder.

Um die Nützlichkeit des Vorschlagtools zu bewerten, wurde der HIQuA in Zengs Studie von einer Gruppe von 213 Personen in Verbindung mit dem Suchdienst Google getestet. Das Web wurde einerseits nach vordefinierten Gesundheitsinformationen, aber auch mit einem durch die Probanden individuell erstellten Abfragebegriff durchsucht.

Bei der Hälfte der Teilnehmer wurden die Abfragevorschläge des HIQuA ausgeblendet, diese Personen erhielten also keine Empfehlungen für die Suche. Die Ergebnisse der beiden Gruppen wurden danach hinsichtlich der Zufriedenheit und der Erfolgsrate verglichen.

Kein statistisch signifikanter Unterschied konnte zwischen beiden Gruppen hinsichtlich des Kriteriums der Benutzer-Zufriedenheit festgestellt werden. Die Gruppe der Personen, die Empfehlungen durch den HIQuA erhielten, war zu 85,2% zufrieden mit den Ergebnissen der Abfrage, die Gruppe, die ohne solche Empfehlungen suchen musste waren dies zu 80,6%.

Ein statistisch signifikanter Unterschied ($p = 0,006$) wurde bei der Frage der Erfolgsrate der Abfragen zwischen beiden Gruppen ermittelt, dabei war die Erfolgsrate definiert als erfolgreich, sobald eines oder mehrere relevante Dokumente in den Top 10 Ergebnislisten zurückgeliefert wurden. Die Gruppe der Personen, die durch HIQuA unterstützt wurde, erzielte 76,0% erfolgreiche Abfragen. Die Gruppe der Personen ohne Empfehlungshilfe durch HIQuA erzielten 65,7% erfolgreiche Abfragen.¹⁰⁴

Die Forscher deuteten das Ergebnis als erfolgreich, obwohl auch darauf hingewiesen wurde, dass hinsichtlich des wesentlichen Kriteriums der Zufriedenheit keine statistisch signifikante Verbesserung durch den Einsatz des HIQuA erzielt worden ist.

Die wesentlichen Erkenntnisse lassen aber den Schluss zu, dass der neue Zugang mit Semantic Web Technologie eine Verbesserung gegenüber den bisherigen, thesaurusbasierten Systemen bietet. Weil die Formulierung einer Abfrage eine

¹⁰² Medline Plus, URL <http://www.nlm.nih.gov/medlineplus/>.

¹⁰³ Vgl. Zeng, Q. T. et al, 'Assisting consumer health information retrieval with query recommendations' (2006) 13(1) *J Am Med Inform Assoc* 80-90, S. 82.

¹⁰⁴ Vgl. Zeng, Q. T. et al, 'Assisting consumer health information retrieval with query recommendations' (2006) 13(1) *J Am Med Inform Assoc* 80-90, S.85-87.

herausfordernde Aufgabe für Informationssuchende ist, wird diesen und ähnlichen Tools in der Zukunft aber mehr Bedeutung zugemessen.¹⁰⁵

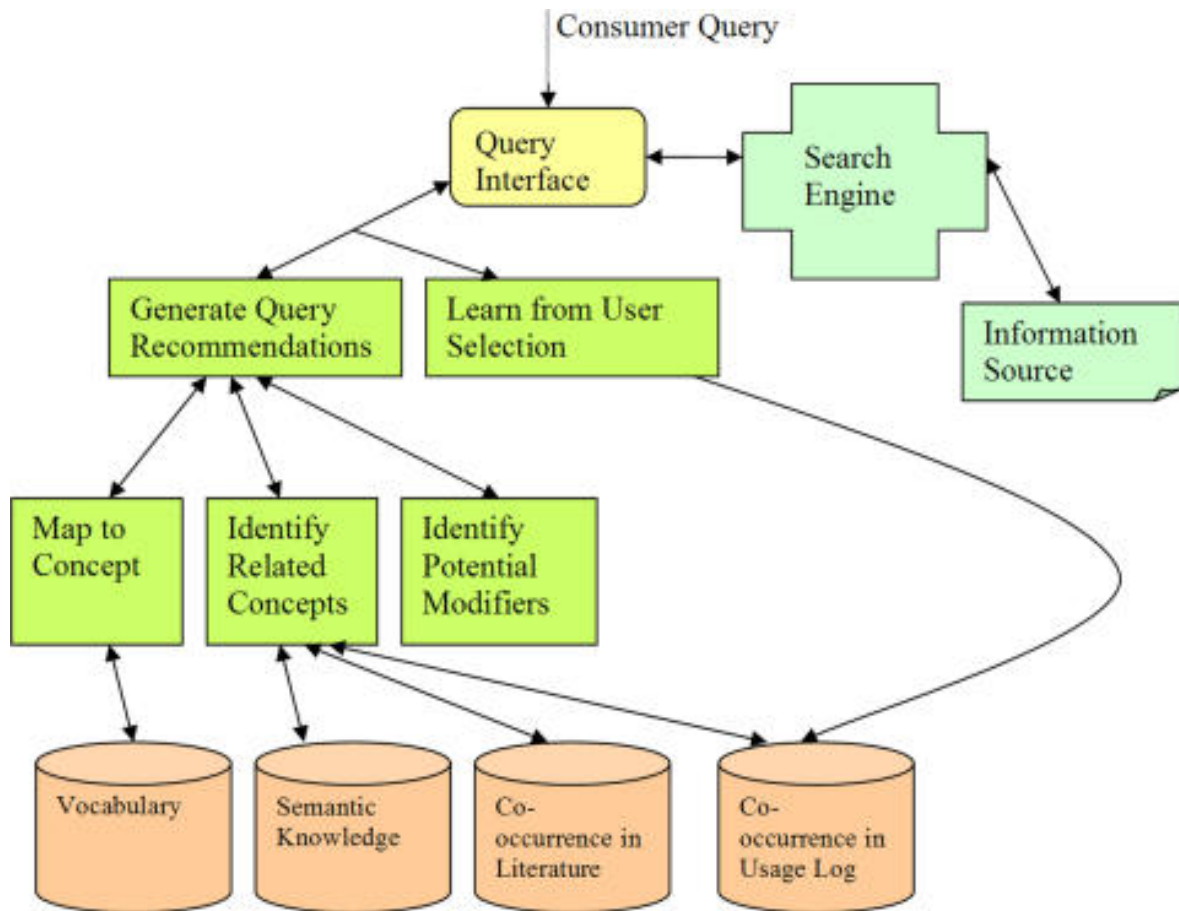


Abbildung 7: Design des HIQuA (nach Zeng et al, 2006)

Es liegen keine Informationen vor, ob der HIQuA das Entwicklungsstadium bereits verlassen hat. Daher sind auch keine weiterführenden Informationen über Verbreitung und Akzeptanz bei den Benutzern verfügbar.

3.5 Wikipedia und Encyclopedia Britannica

Der bisher meist beachtete Vergleich zwischen dem autorenbasierten System der Encyclopedia Britannica und dem kollaborativen System der Wikipedia wurde von Jim Giles¹⁰⁶ in *Nature* dargestellt. Dieser Artikel hat heftige Kontroversen mit den Verantwortlichen der Encyclopedia Britannica hervorgerufen, die *Nature* schlicht

¹⁰⁵ Vgl. Zeng, Q. T. et al, 'Assisting consumer health information retrieval with query recommendations' (2006) 13(1) *J Am Med Inform Assoc* 80-90, S.88.

¹⁰⁶ Vgl. Giles, Jim, 'Internet encyclopaedias go head to head' (2005) 438 *Nature* 900-901.

Unwissenschaftlichkeit vorwarfen. *Nature* konnte jedoch alle nötigen Nachweise erbringen, die die ordnungsgemäße Durchführung bestätigten.

Die 2005 veröffentlichte Studie wurde aufgrund der außergewöhnlichen Entwicklung, die Wikipedia genommen hatte, durchgeführt. Das rapide Wachstum der Artikel war ein derart auffälliges Phänomen, dass die Forscher die Frage nach der Verlässlichkeit der Information zu stellen begannen.

Die Artikel in allen Sprachausprägungen von Wikipedia hatten damals eine Zahl von 3,7 Millionen erreicht. Täglich wurden zur englischsprachigen Version 1.500 Artikel hinzugefügt. Der Alexa Web-Ranking-Dienst führte *wikipedia.org* an 37. Stelle der Top 100 Websites (aktuell Platz 10¹⁰⁷).

Das Kernergebnis dieses Vergleichs ist, dass beide Online-Enzyklopädien gute Informationsquellen, beide aber nicht gänzlich ohne Fehler sind, und, dass die Fehlerhäufigkeit weder in der einen noch in der anderen Enzyklopädie signifikant unterschiedlich ist.

Die Studie wählte Artikel aus einer Vielzahl von Wissenschaftsdisziplinen aus den Webseiten beider Enzyklopädien und ließ diese in anonymisierter Form von Experten begutachten. Den Experten war die Herkunft der Artikel nicht bekannt. Jeder Experte bewertete einen einzelnen Gegenstand aus beiden Enzyklopädien. Von 50 ausgesendeten Anfragen wurden 42 zurückgesendet und für die weitere Begutachtung durch das Team von *Nature* ausgewertet.

Nur insgesamt 8 als erheblich eingestufte Fehler, wie etwa die falsche Interpretation von wichtigen Konzepten, wurden gefunden, je 4 pro Enzyklopädie.

Insgesamt wurden von den Experten aber auch 285 kleinere Fehlern (Wikipedia 162, Britannica 123), gefunden. Bei diesen handelte es sich etwa um sachlich unkorrekte Einträge oder aber auch um Auslassungen wichtiger Fakten.

In der Arbeit wurden ebenso problematische Seiten von Wikipedia angesprochen, die zu der Zeit der Studienerstellung allgemein für kontroversielle Diskussionen gesorgt hatten. Damals aktuelle Fälle, in denen Artikel in Wikipedia bewusst manipuliert wurden, sind vom Autor als „*the exception rather than the rule*“¹⁰⁸ bezeichnet worden.

Der grundsätzliche Unterschied zwischen einem System mit überwachenden Editoren und einem aus einer Vielzahl von gleichberechtigten Autoren im kollaborativen System

¹⁰⁷ Vgl. o.V., *Alexa Top 500 Web Sites* (2007) Alexa Internet, Inc.

<http://www.alexa.com/data/details/traffic_details?url=wikipedia.org> am 10.05.2007 .

¹⁰⁸ Giles, Jim, 'Internet encyclopaedias go head to head' (2005) 438 *Nature* 900-901.

der Wikipedia ändert nach den Ergebnissen dieser Studie nichts an der Qualität der Informationen.¹⁰⁹

3.6 Ergebnisse der Feldanalyse

Die Abfragen auf de.wikipedia.org und google.at und die nachfolgenden Analysen wurden am 26. und 27.2.2007 durchgeführt.

Tabelle 2 zeigt den genauen Wortlaut der eingegebenen Abfragebegriffe (N=30) der durch Zufallsprinzip ermittelten Krankheitsbegriffe, die von Wikipedia zurück gelieferten Artikelbezeichnungen, die Anzahl der Versionen jedes vorhandenen Artikels und das Datum der Entstehung. Unter Relevanz ist die Anzahl der weiteren erforderlichen Eingaben (Mausklicks) zu verstehen, um zum relevanten Artikel zu gelangen, falls dieser nicht direkt als Rückantwort geliefert wurde. 100% bedeutet, der Abfragestring führt direkt auf einen entsprechenden Artikel. Jede weitere Eingabe (Mausklick) führt zu einer Halbierung des Wertes. Als Beispiel 50% dann, wenn für das Erreichen des Artikels eine weitere Eingabe notwendig war, weil die Abfrage mehrere mögliche Artikel in einer Auswahlliste zurücklieferte.

3.6.1 Wikipedia Abfrage

28 Abfragebegriffe zeigten direkt auf entsprechende Artikel in der Wikipedia, zwei weitere wurden über eine Auswahlliste, die alle relevanten Zwischenergebnisse anzeigte, erreicht.

Die Top-5 Artikel mit der größten Anzahl an Versionen sind Cholesterin (1160), Depression (1141), Zuckerkrankheit (1079), Adipositas (807) und Multiple Sklerose (685). Die Artikel mit den wenigsten Versionen sind Schlaganfall (20), Myokarditis (51), COPD (78), Toxoplasmose (83) und Scharlach (103) bzw. Arthritis (103).

Der älteste Artikel aus der Auswahl der 30 Krankheitsbilder in der deutschen Wikipedia ist Multiple Sklerose, der jüngste Schlaganfall.

Nur zwei Artikel existieren kürzer als 1.000 Tage, der Mittelwert liegt bei 1.265 Tagen, der am längsten existierende Artikel ist 1.984 Tage alt.

¹⁰⁹ Vgl. Giles, Jim, 'Internet encyclopaedias go head to head' (2005) 438 *Nature* 900-901.

Abfragebegriff	Wikipedia Artikel	Relevanz (%)	Anzahl der Versionen	Erstellungsdatum
Schlaganfall	Schlaganfall	100	20	19.12.2006
Myokarditis	Myokarditis	100	51	31.08.2004
COPD	COPD	100	78	20.10.2003
Toxoplasmose	Toxoplasmose	100	83	08.04.2004
Arthritis	Arthritis	100	103	29.10.2003
Scharlach	Scharlach (Krankheit)	100	103	06.11.2003
Koronare Herzerkrankung	Koronare Herzkrankheit	50	107	29.11.2002
Aneurysma	Aneurysma	100	117	09.03.2004
Pankreatitis	Pankreatitis	100	119	07.03.2004
Anämie	Anämie	100	128	23.11.2003
Leberzirrhose	Leberzirrhose	100	141	04.11.2002
Colitis ulcerosa	Colitis ulcerosa	100	178	07.02.2004
Osteoporose	Osteoporose	100	238	02.02.2004
Lungenkrebs	Lungenkrebs	100	241	09.02.2004
Morbus Hodgkin	Morbus Hodgkin	100	257	04.03.2004
Rheumatismus	Rheuma	100	290	12.01.2004
Asthma	Asthma	100	297	04.02.2003
Poliomyelitis	Poliomyelitis	100	305	19.11.2003
Pfeiffersches Drüsenfieber	Pfeiffer-Drüsenfieber	100	327	06.03.2004
Bluthochdruck	Bluthochdruck	100	363	18.09.2002
Brustkrebs	Brustkrebs	100	386	01.08.2002
Alzheimer	Alzheimer-Krankheit	100	486	01.09.2003
Herzinfarkt	Herzinfarkt	100	530	02.07.2002
Gastritis	Gastritis	100	634	30.10.2003
Hepatitis	Hepatitis	100	646	07.01.2003
Multiple Sklerose	Multiple Sklerose	100	685	24.09.2001
Adipositas	Adipositas	100	807	03.11.2003
Zuckerkrankheit	Diabetes mellitus	100	1079	26.10.2003
Depression	Depression	100	1141	03.11.2003
Cholesterin	Cholesterin	50	1160	25.07.2002
N=30, 30 erfolgreiche Rückantworten				

Tabelle 2: Ergebnisse der Wikipedia Abfrage (Autor)

3.6.2 Kontrollprobe

Von der Gruppe der verbliebenen 30 Krankheiten wurde am 1.3. und am 2.3.2007 ebenfalls eine Feldanalyse durchgeführt.

28 Abfragebegriffe (Tabelle 3) zeigten direkt auf korrespondierende Artikel in Wikipedia, ein weiterer wurde über eine Auswahlliste, die alle relevanten Zwischenergebnisse anzeigte, erreicht, zu einem Begriff existierte kein Artikel in der Wikipedia.

Die Top-5 Artikel mit der größten Anzahl an Versionen sind Epilepsie (920), HIV (830), Morbus Chron (466), Pneumonie (296), Tetanus (267).

Die Artikel mit den wenigsten Versionen sind Mykose (10), Purpura (14), Phäochromozytom (27), Divertikulitis (31) und Raynaud-Syndrom (41).

Der älteste Artikel ist HIV, der jüngste Phäochromozytom. Das mittlere Alter aller Artikel liegt bei 1090 Tagen.

Abfragebegriff	Wikipedia Artikel	Relevanz (%)	Anzahl der Versionen	Erstellungsdatum
Mykose	Mykose	100	10	20.05.2004
Purpura	Purpura	100	14	24.06.2005
Phäochromozytom	Phäochromozytom	100	27	09.08.2005
Divertikulitis	Divertikulitis	100	31	07.03.2005
Morbus Raynaud	Raynaud-Syndrom	100	41	31.03.2004
Zytomegalie	Zytomegalie	100	43	25.04.2005
Sjögren Syndrom	Sjögren-Syndrom	100	46	05.11.2004
Lymphom	Malignes Lymphom	50	51	08.10.2004
EPH-Gestose	Präeklampsie	100	53	29.06.2004
Struma	Struma	100	56	23.05.2005
Trigeminusneuralgie	Trigeminusneuralgie	100	79	18.11.2004
Amyloidose	Amyloidose	100	80	27.10.2004
Tularämie	Tularämie	100	94	23.06.2004
Madenwurm	Madenwurm	100	96	10.03.2004
Myasthenia gravis	Myasthenia gravis pseud.	100	97	24.02.2003
Turner Syndrom	Turner Syndrom	100	103	05.03.2004
Chronische Polyarthrit	rheumatoide Arthritis	100	146	24.07.2004
Appendizitis	Appendizitis	100	146	29.09.2003
Gicht	Gicht	100	146	12.08.2003
Herpes Zoster	Herpes Zoster	100	175	02.05.2004
Vorhofflimmern	Vorhofflimmern	100	186	31.12.2002
Thrombose	Thrombose	100	201	08.11.2003
Skoliose	Skoliose	100	204	21.02.2004
Herzinsuffizienz	Herzinsuffizienz	100	247	15.09.2002
Tetanus	Tetanus	100	267	17.12.2003
Pneumonie	Lungenentzündung	100	296	03.10.2003
Morbus Crohn	Morbus Crohn	100	446	07.02.2004
HIV	HIV	100	830	14.09.2002
Epilepsie	Epilepsie	100	920	10.11.2002
Cholecystitis	n/a	n/a	n/a	n/a
N=30, 29 erfolgreiche Rückantworten				

Tabelle 3: Ergebnisse der Wikipedia Kontrollprobe (Autor)

Vergleicht man das Entstehungsdatum der jeweiligen Artikel aus der Probe der Feldanalyse (N=30) mit der Kontrollprobe (N=29) lässt sich erkennen, dass in beiden Proben bis zum Jahr 2003 63%, respektive 83% aller Artikel bereits angelegt worden

sind. Bis zum Ende des Jahres 2004 waren in beiden Proben unabhängig voneinander bereits 97% aller Artikel angelegt (Abb.8).

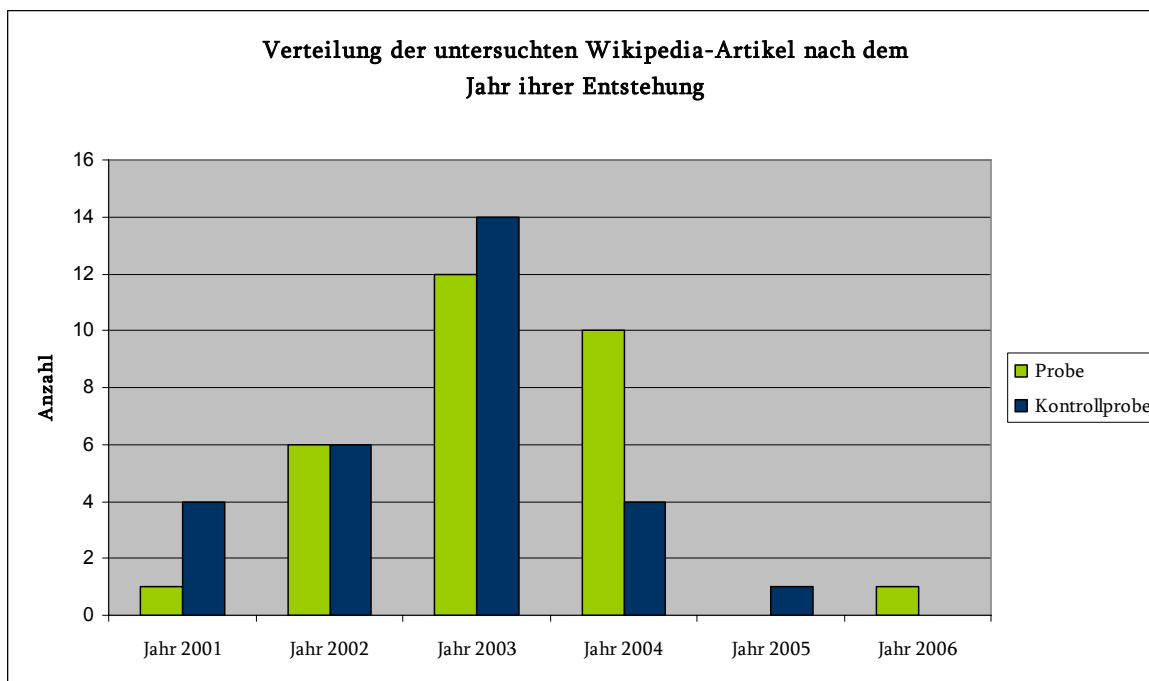


Abbildung 8: Entstehungszeit der Wikipedia-Artikel (Autor)

3.6.3 Google Abfrage

Die Abfrage mit den Begriffen aus der primären Feldanalyse sowie mit den Begriffen aus der Kontrollprobe wurde in der österreichischen Version des Suchdienstes Google durchgeführt. Es wurden nur die Ergebnisse der offiziellen Reihung (Ranking) des Google PageRank in Betracht gezogen, bezahlte Anzeigen wurden nicht berücksichtigt.

Google verwendet für die Präsentation der Suchergebnisse den so genannten PageRank-Algorithmus, der besonderen Wert auf die Relevanz einer Seite legt.

Der Algorithmus basiert, soweit öffentlich bekannt, auf der Annahme, dass die Anzahl der Links, die auf bestimmte Internetseiten zeigen, ein Kriterium für die Wichtigkeit ist. Häufige Links auf eine Seite ist ein Kennzeichen einer wichtigen Seite. Der PageRank auf die Seite A setzt voraus, dass deren Verlinkungseigenschaften bekannt sind.

Die vereinfachte Formel des PageRank-Algorithmus¹¹⁰ ist im weiteren Sinne ebenfalls eine Web 2.0 Anwendung, da sie auf die kollektive Intelligenz des Web¹¹¹ setzt.

¹¹⁰ Vgl. Brin, Sergey / Lawrence Page, 'The anatomy of a large-scale hypertextual {Web} search engine' (1998) 30 *Computer Networks and ISDN Systems* 107-117 S. 110.

¹¹¹ Vgl. 3.2.1.3, S. 35.

Sie lautet:

Sind $T_1 \dots T_n$ die Seiten die auf A verlinken
und $C(T_i)$ die Anzahl der Links die aus einer Seite T_i herausgehen, dann gilt:

$$PR(A) = (1-d) + d(PR(T_1)/C(T_1) + \dots + PR(T_n)/C(T_n))$$

d:

Dämpfungsfaktor

Die aus einer Google-Abfrage zurück gelieferten Werte werden vom Suchdienst, im Sinne des PageRank, nach absteigender Relevanz sortiert.

3.6.3.1 Ergebnisse der primären Krankheitsbegriffe

Abbildung 9 zeigt die graphische Verteilung der Webressourcen, die für jeden Abfragebegriff den Spitzenplatz im Google-Ranking eingenommen haben.

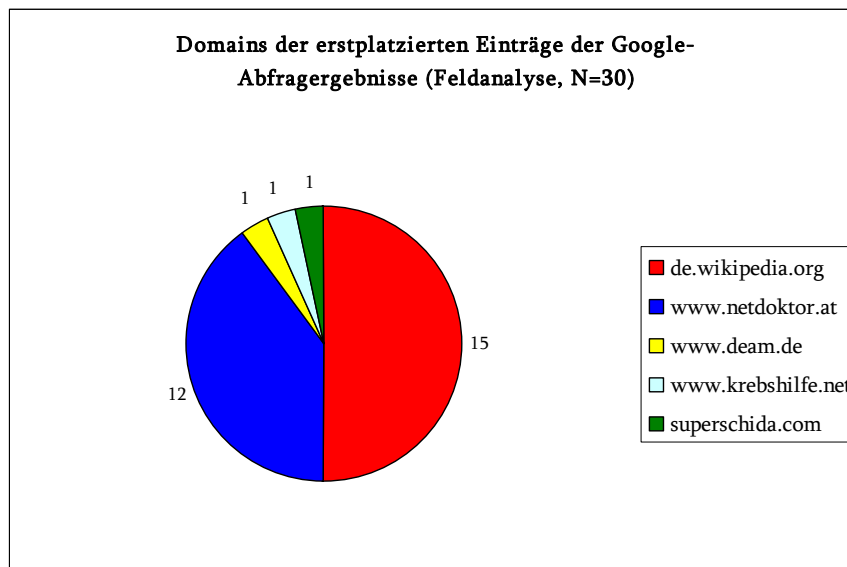


Abbildung 9: Verteilung des Rankings, primäre Krankheitsbegriffe (Autor)

Die Abfrage der 30 primär verwendeten Krankheitsbegriffe brachte im Detail folgendes Ergebnis: 15-mal war die URL des entsprechenden Wikipedia-Artikels erstgereiht im Listing der Suchmaschine.

7-mal war der entsprechende Artikel in Wikipedia der zweit gereichte Eintrag, wobei in 5 solchen Fällen eine entsprechende URL auf www.netdokter.at den ersten Platz eingenommen hat. Details sind der Tabelle 4 zu entnehmen.

Auffällig ist, dass www.netdokter.at viele Spitzenergebnisse lieferte, insgesamt nämlich 12 erstgereichte Artikel. Ein Wikipedia-Artikel, Scharlach, war nicht in den Top-Ten Platzierungen vertreten, er rangiert auf Platz 22.

Der Mittelwert aller Wikipedia-Rankings aus den 30 Abfragebegriffen ergibt 2,6, der Median 2.

Abfragebegriff	Wikipedia Rank	Andere
Schlaganfall	1	-
Arthritis	1	-
Koronare Herzerkrankung	1	-
Aneurysma	1	-
Pankreatitis	1	-
Leberzirrhose	1	-
Osteoporose	1	-
Asthma	1	-
Alzheimer	1	-
Hepatitis	1	-
Multiple Sklerose	1	-
Adipositas	1	-
Depression	1	-
Cholesterin	1	-
Poliomyelitis	1	-
Rheumatismus	2	www.deam.de
Brustkrebs	2	www.krebshilfe.net
Anämie	2	www.netdokter.at
Lungenkrebs	2	www.netdokter.at
Morbus Hodgkin	2	www.netdokter.at
Herzinfarkt	2	www.netdokter.at
Gastritis	2	www.netdokter.at
COPD	3	www.netdokter.at
Colitis ulcerosa	3	superschida.com
Myokarditis	4	www.netdokter.at
Toxoplasmose	4	www.netdokter.at
Pfeiffersches Drüsenfieber	4	www.netdokter.at
Zuckerkrankheit	4	www.netdokter.at
Bluthochdruck	5	www.netdokter.at
Scharlach	22	www.netdokter.at

Tabelle 4: Google Rangliste der Primärabfrage (Autor)

3.6.3.2 Ergebnis der Kontrollprobe

Abbildung 10 zeigt die graphische Verteilung der Webressourcen, die für jeden Abfragebegriff der Kontrollprobe den Spitzenplatz im Google-Ranking eingenommen haben.

Die Abfrage der 30 verbliebenen Krankheitsbegriffe aus der Kontrollprobe (Details siehe Tabelle 5) brachte folgendes Ergebnis: 14-mal war die URL des entsprechenden Artikels in der Wikipedia erstgereiht. 7-mal war es eine URL des österreichischen Anbieters netdoktor.at.

Zweimal war eine URL des Consilium des Allgemeinen Krankenhauses der Stadt Wien, das seine Informationen aber nur medizinisch geschultem Personal empfiehlt, erstplatziert. Die weiteren Erstplatzierungen zeigten überwiegend auf URLs von Selbsthilfegruppen.

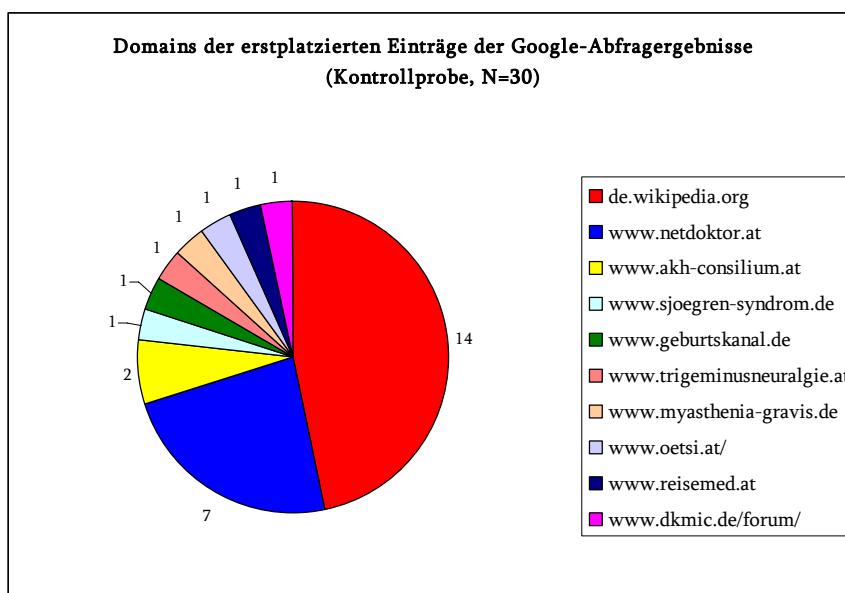


Abbildung 10: Verteilung des Rankings, sekundäre Krankheitsbegriffe (Autor)

Abfragebegriff	Wikipedia Rank	Andere
Mykose	1	-
Purpura	1	-
Morbus Raynaud	1	-
Amyloidose	1	-
Tularämie	1	-
Madenwurm	1	-
Appendizitis	1	-
Herpes Zoster	1	-
Vorhofflimmern	1	-
Thrombose	1	-
Skoliose	1	-
Morbus Crohn	1	-
HIV	1	-
Epilepsie	1	-
Myasthenia gravis	2	www.myasthenia-gravis.de
Turner Syndrom	2	www.oetsi.at/
Gicht	2	www.netdokter.at
Herzinsuffizienz	2	www.netdokter.at
Tetanus	2	www.reisemed.at
Pneumonie	2	www.netdokter.at
Lymphom	3	www.netdokter.at
Divertikulitis	4	www.netdokter.at
Zytomegalie	4	www.akh-consilium.at
Sjögren Syndrom	4	www.sjoegren-syndrom.de
EPH-Gestose	4	www.geburtskanal.de
Trigeminusneuralgie	5	www.trigeminusneuralgie.at
Struma	7	www.netdokter.at
Chronische Polyarthritits	7	www.netdokter.at
Phäochromocytom	n/a	medd.klinikum.uni-muenster.de
Cholecystitis	n/a	www.dkmic.de/forum/

Tabelle 5: Google Rangliste der Kontrollprobe (Autor)

Der Mittelwert aller Wikipedia-Rankings aus den 30 Abfragebegriffen der Kontrollprobe ergibt 2,3, der Median 2.

3.7 Ergebnisse der medizinischen Analyse

Die Überprüfung des medizinischen Inhaltes der 30 Artikel in Wikipedia aus der primären Zufallsprobe wurde nach den Kriterien (1) Umfang im Sinne von Übersichtlichkeit und Vollständigkeit, (2) Richtigkeit und (3) Verständlichkeit für Laien geprüft. Die Bewertung wurde von einem Facharzt durchgeführt. Zur Beurteilung wurde ein numerisches Bewertungsschema von 1-5 verwendet, wobei 1 sehr gut, 2 gut, 3 durchschnittlich, 4 akzeptabel und 5 nicht akzeptabel entspricht.

Nr.	Wiki Artikel Titel	Umfang	Richtigkeit	Verständlichkeit
1	Adipositas	1	1	1
2	Herzinfarkt	1	1	1
3	Asthma	1	2	1
4	Lungenkrebs	3	1	2
5	COPD	2	1	1
6	Brustkrebs	1	1	1
7	Koronare Herzkrankheit	3	2	1
8	Cholesterin	4	2	2
9	Diabetes mellitus	1	1	1
10	Bluthochdruck	1	1	1
11	Depression	1	1	2
12	Arthritis	2	1	1
13	Osteoporose	1	1	1
14	Anämie	4	2	4
15	Multiple Sklerose	1	2	1
16	Alzheimer-Krankheit	1	1	1
17	Hepatitis	3	1	2
18	Gastritis	3	2	1
19	Rheuma	3	3	2
20	Schlaganfall	3	3	1
21	Pfeiffer-Drüsenfieber	1	1	1
22	Scharlach (Krankheit)	2	2	1
23	Toxoplasmose	4	1	3
24	Aneurysma	4	1	5
25	Morbus Hodgkin	1	1	1
26	Myokarditis	2	1	1
27	Poliomyelitis	1	1	1
28	Leberzirrhose	1	1	1
29	Colitis ulcerosa	1	1	2
30	Pankreatitis	2	1	2

Tabelle 6: Ergebnisse der medizinischen Fachbewertung (Autor)

Das Resultat der Analyse ergibt in Summe eine gute bis sehr gute Bewertung mit einigen wenigen negativen Ausreißern, die die Qualität aber nicht wesentlich beeinträchtigen. In den folgenden Unterkapiteln werden die Ergebnisse der medizinischen Analyse detailliert dargestellt.

3.7.1 Umfang

Hinsichtlich des Umfangs (Abbildung 11) wurde keine schlechtere Bewertung als 4 abgegeben, wobei bei drei der mit 4 bewerteten Artikel (Cholesterin, Anämie, Aneurysma) die zusätzliche Information gegeben wurde, diese Artikel wären in hohem Maß zu umfangreich und daher ungeeignet für die Verwendung durch Laien.

Einer dieser vier Artikel (Toxoplasmose) wurde als unvollständig bewertet. Die Durchschnittsbewertung (Mittelwert) aller 30 Artikel hinsichtlich des Umfangs lag bei 2,0. 50% aller Artikel wurde mit sehr gut bewertet.

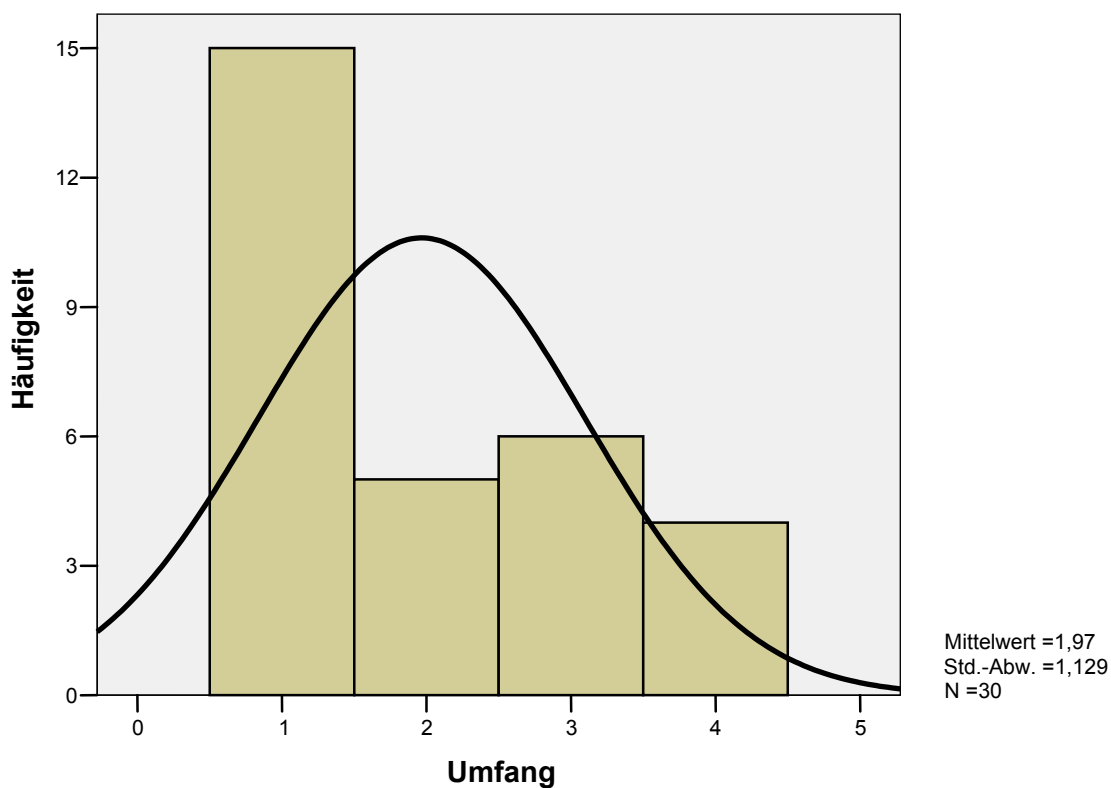


Abbildung 11: Verteilung von Umfang (Autor)

3.7.2 Richtigkeit

Die medizinische Richtigkeit (Abbildung 12) wurde durchschnittlich (Mittelwert) mit 1,4 bewertet, der schlechteste Wert ist hier zweimalig mit 3 gekennzeichnet, nämlich Rheuma und Schlaganfall. 21-mal wurde die Note 1 vergeben, das entspricht einem prozentuellen Anteil von 70.

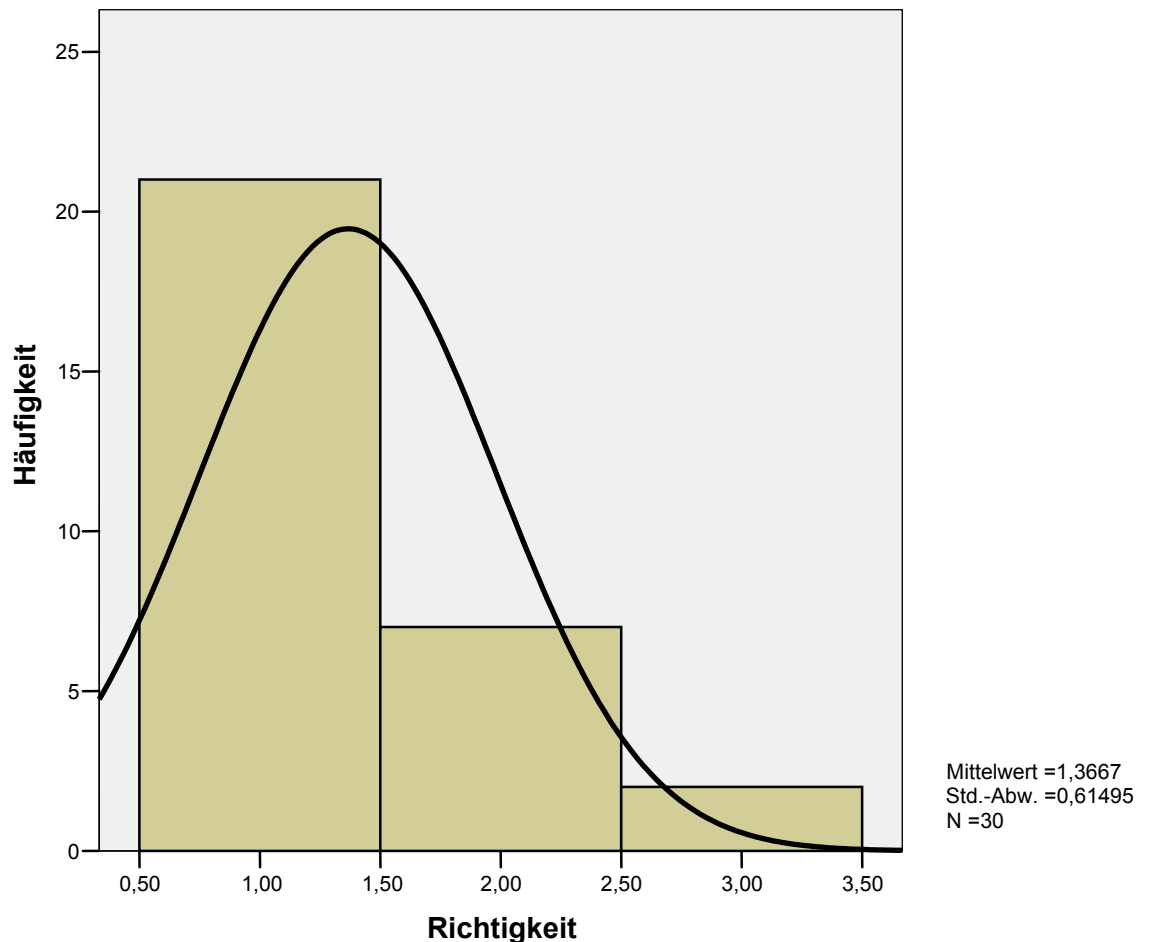


Abbildung 12: Verteilung von Richtigkeit (Autor)

3.7.3 Verständlichkeit für Laien

Die Verständlichkeit für Laien (Abbildung 13) wurde durchschnittlich (Mittelwert) mit 1,5 bewertet. Ein Artikel (Aneurysma) wurde mit 5, nicht akzeptabel bewertet. Die Begründung dafür war, dass der Artikel sowohl durch seinen großen Umfang und seine Detaillierung sogar für medizinisches Fachpersonal ungeeignet ist. Ein weiterer Artikel (Anämie) wurde mit der Note 4 bewertet, weil er aus der Sicht des Mediziners keine vollständige Information für Laien bot. Ein Artikel (Toxoplasmose) wurde mit 3

bewertet, die Begründung dafür wurde mit einem zu bedrohlichen Schreibstil angegeben.

Alle anderen Artikel wurden hinsichtlich der Verständlichkeit für Laien als gut (7 Artikel) und sehr gut (20 Artikel) bewertet. Der Anteil aller mit sehr gut bewerteten Artikel lag bei 66,7%.

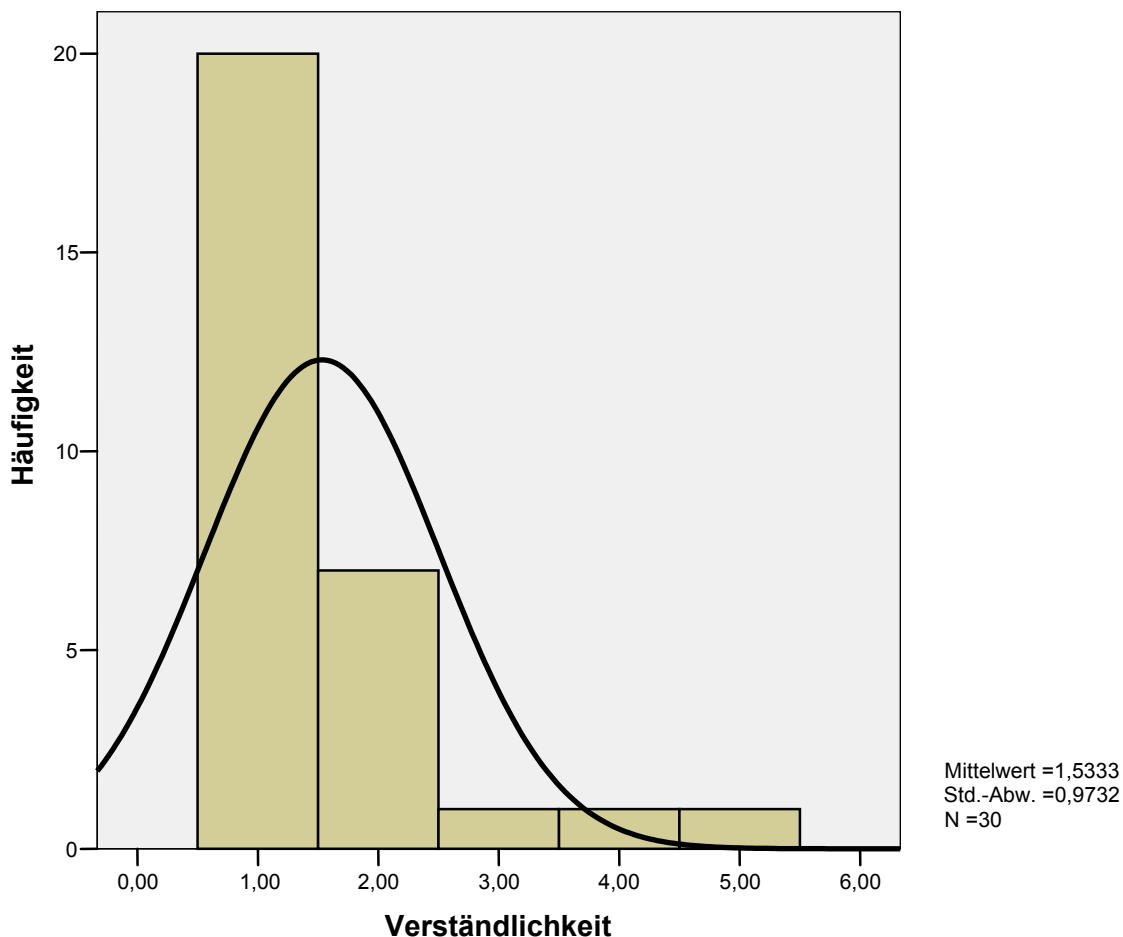


Abbildung 13: Verteilung von Verständlichkeit (Autor)

3.7.4 Gesamtbetrachtung der medizinischen Qualität

Insgesamt 8 Artikel wurden in jeden einzelnen Kategorien mit der Note 1 bewertet. Die schlechteste Durchschnittsbewertung mit je 3,3 erhielten die Artikel Anämie und Aneurysma. Der Mittelwert aller Bewertung der Artikel in allen Kategorien ergab den Wert 1,6.

Abbildung 14 zeigt die Boxplots der drei Kriterien Umfang, Richtigkeit und Verständlichkeit aus allen Kategorien, konkret Minimum, Quartile 1, Median, Quartile 3, Maximum).

Die beiden Ausreißer Anämie (Datenbeschriftung 14 in Abb. 14) und Aneurysma (Datenbeschriftung 24 in Abb. 14) wurden auch in der Kategorie Umfang schlecht bewertet.

Alle drei Kriterien sind normalverteilt, sodass Aussagen über die Wahrscheinlichkeit getroffen werden können. Mit einer 99% Wahrscheinlichkeit (Confidence Interval 99%), wird bei einer weiteren Probe bezüglich des Kriterium Umfang der Mittelwert zwischen 1,40 und 2,53 liegen. Mit einer 99% Wahrscheinlichkeit liegt der Mittelwert einer erneuten Untersuchung bezüglich des Kriteriums Richtigkeit in einem Bereich von 1,1 bis 1,68. Eine neuerliche Untersuchung des Kriteriums Verständlichkeit würde mit gleich hoher Wahrscheinlichkeit zu einem Mittelwert von 1,04 bis 2,02 führen.

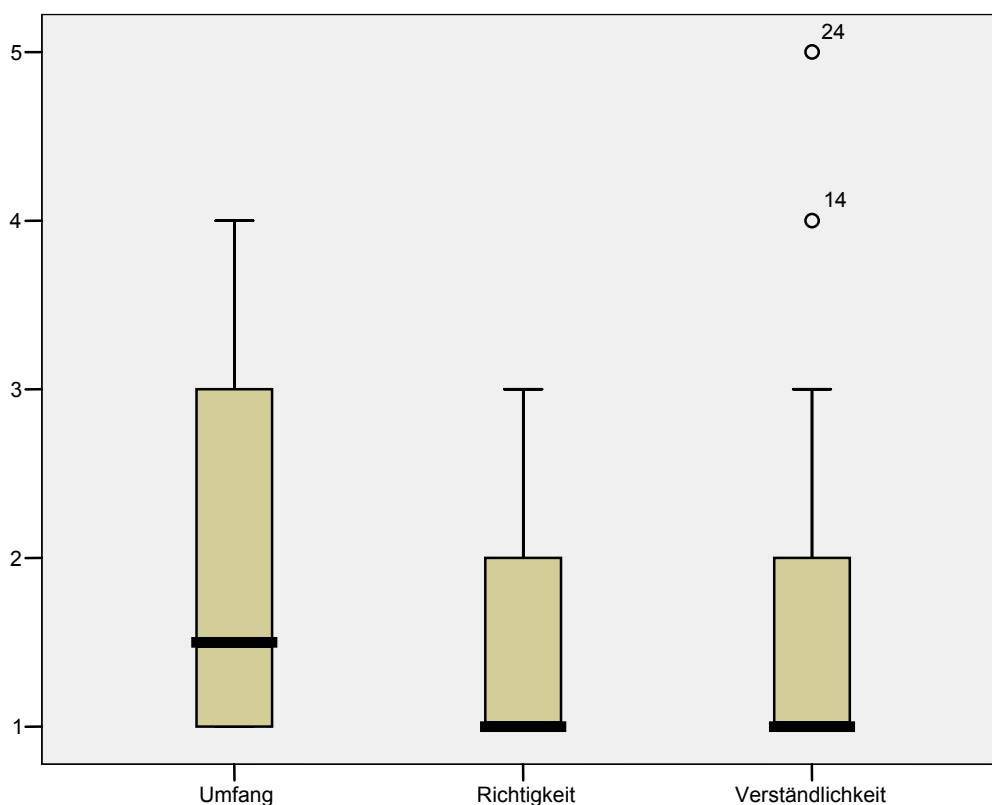


Abbildung 14: Boxplots aller drei Wertungskriterien (Autor)

3.7.5 Kennzeichnungspraxis medizinischer Information in der Wikipedia

Alle untersuchten medizinischen Artikel in der Wikipedia sind besonders als solche gekennzeichnet. Ein weißes Kreuz auf grünem Grund (Abbildung 15) symbolisiert am Ende eines Artikels, dass es sich hierbei um ein Gesundheitsthema handelt.



Abbildung 15: Symbol zur Kennzeichnung von Gesundheitsartikel (Wikipedia)

Das Symbol ist mit den entsprechenden Informationen, die auf die Gefahren der einseitigen Verwendung von Gesundheitsinformation aus dem World Wide Web hinweist, verlinkt (Abbildung 16). Abbildung 16 zeigt auch den Grundaufbau von Wikipedia mit der Basisnavigation (linker Bildrand), dem Artikelregister (hier Projektseite), dem Diskussionsregister, dem Quelltextregister sowie dem Versions- und Autorenregister eines Artikels.



Abbildung 16: Hinweis auf Gesundheitsinformation bei Wikipedia (Wikipedia)

3.8 Zutreffen der Qualitätskriterien auf Wikipedia-Artikel

Die Daten bzw. die gesundheitsbezogenen Artikel der Feldanalyse aus Wikipedia wurden mit den in der Literatur genannten Qualitätskriterien der *Codes of Conduct* verglichen. Für den Vergleich wurde der HON Code (Europa) und die von der amerikanischen Gesundheitsbehörde im Rahmen der Gesundheitsinitiative Healthy People 2010 veröffentlichten Kriterien verwendet, sofern diese verschieden von den Kriterien der HON waren.

In Tabelle 7 werden die Kriterien kurz beschrieben. In der Spalte mit dem Titel Wikipedia Fulfillment wird angegeben, zu welchem Grad Wikipedia die Bedingungen aufgrund der vorliegenden Forschungsergebnisse erfüllen kann.

Die Qualitätskriterien, die für die Darstellung von Gesundheitsinformationen im World Wide Web aufgestellt wurden, werden von Wikipedia nahezu alle im vollen Umfang erfüllt.

Das Kriterium *Authority* wird von Wikipedia sicherlich nur teilweise erfüllt, da es hinsichtlich der Autorenschaft keine einschränkenden Regeln gibt. Ein Arzt kann hier genauso publizieren wie ein Laie.

Das Kriterium *Transparency* wird von Wikipedia erfüllt, da Autoren zwar üblicherweise lediglich mit ihrer Online-Identität anonym, aber über das System der Wikipedia jederzeit erreichbar sind.

Alle anderen Kriterien werden von Wikipedia zur Gänze erfüllt.

Kriterium	Herkunft	Beschreibung	Wikipedia Fulfillment
Authority	HON	Wird medizinische Information durch einen Mediziner gegeben?	Zum großen Teil ja, aber auch betroffene Laien oder Angehörige können zur Erweiterung beitragen
Complementarity	HON	Information muss unterstützen, nicht die Arzt-Patientenbeziehung ersetzen	Wikipedia weist bei Gesundheitsartikeln ausdrücklich darauf hin, dass Information nur unterstützend ist
Privacy	HON	Wie werden personenbezogene Daten - die durch eine Website gesammelt werden können - verwendet?	Es werden kein persönlichen Daten, außer Benutzeranmeldedaten, gesammelt
Attribution	HON	Gewissenhafte und genaue Referenzierung auf die Informationsquellen	Wikipedia hält vom Prinzip der Enzyklopädie her darauf bedacht, dass Autoren verlässlich auf Quellen referenzieren
Justifiability	HON	Die Wirksamkeit und der Nutzen eine Therapie oder eines Produktes werden ausgewogen dargestellt	Das oberste Prinzip bei der Bereitstellung von Information ist die ausgewogene Darstellung
Transparency	HON	gut zugänglich Darstellung der Identität von Editoren und Verantwortlichen, Feedbackmöglichkeit	Jeder Autor ist über die Diskussionsseite jederzeit erreichbar, da es sich um Nicknames handelt, ist die wahre Identität eines Autors nicht feststellbar
Financial disclosure	HON	Klare Benennung aller Sponsoren, die finanziell zu Entstehung und zum Betrieb beitragen	Wikipedia ist ein Non-Profit-Organisation und auf Spenden angewiesen.
Sponsorship	HON	Klare Trennung des redaktionellen Inhalts von Werbeinhalt	Wikipedia verzichtet gänzlich auf Werbeeinschaltungen
Aktualität, Update	HHS	Wie wird der Content upgedatet?	Die Versionsliste gibt Auskunft über Art und Datum der Änderung, Versionen können miteinander verglichen werden, z.B. die aktuelle mit einer früheren
Absicht und Zielrichtung	HHS	Welche Zielrichtung verfolgt der Informationsanbieter?	Die Zielrichtung der Wikipedia ist die freie Weitergabe von Wissen

Tabelle 7: Zutreffen der Qualitätskriterien auf Wikipedia (Autor)

4 Diskussion

4.1 Health Information Retrieval

Es ist davon auszugehen, dass auch im deutschsprachigen Raum ein überwiegender Anteil von Nutzern von Gesundheitsinformationen im World Wide Web die Suche über die großen Suchmaschinen Google und Yahoo beginnt. Dieser Nachweis wurde für die USA¹¹² und für einige europäische Länder¹¹³ erbracht und auch die Ergebnisse der Feldanalyse dieser Arbeit legen diesen Schluss nahe.

Für Europa gibt es keine umfassenden Statistiken und Daten in der Art des amerikanischen Online Health Search Reports¹¹⁴, die sich mit spezifischen, groß angelegten Befragungen zum Thema Gesundheitsinformationen im Web befassen.

Während in den USA eine zentrale Einrichtung solche Befragungen über alle Bundesstaaten erhebt, werden die Daten in Europa von den einzelnen statistischen Ämtern der Mitgliedsländer an Eurostat weitergegeben. Die Befragungsbasis der einzelnen Mitgliedsländer ist aber nicht einheitlich geregelt, und es wurden bisher keine spezifischen Fragen über das Verhalten bei der Beschaffung von Gesundheitsinformationen gestellt.

Der Vergleich der Daten zeigt, dass die Nachfrage nach Gesundheitsinformationen im Web in Europa und auch in Österreich ähnlich hoch ist, wie in den USA.

Neben den klassischen Suchmaschinen werden in Zukunft auch spezialisierte, vertikale Suchdienste verstärkt Angebote für Health-Seeker bereitstellen. Vertikale Suchmaschinen, spezialisiert auf Healthcare Information sind bereits verfügbar und können neue Impulse für die Erhöhung der Glaubwürdigkeit setzen¹¹⁵.

Eine Vertikale Suchmaschine kann zwar ganz gezielt die Suche auf Gesundheitswebseiten einschränken und damit treffgenauere Informationen liefern, sie ist dabei aber der direkten Kontrolle des Betreibers ausgesetzt.

¹¹² Vgl. Fox, Susanna *Online Health Search 2006* (2006)

<http://www.pewinternet.org/pdfs/PIP_Online_Health_2006.pdf> am 03.03.2007, S. 5.

¹¹³ Vgl. Kap. 3.1.2. S. 28.

¹¹⁴ Vgl. Fox, Susanna *Online Health Search 2006* (2006)

<http://www.pewinternet.org/pdfs/PIP_Online_Health_2006.pdf> am 03.03.2007.

¹¹⁵ Vgl. 3.3.6, S.46.

Es besteht die Gefahr, dass dadurch die Darstellung von ausgewogenen und vollständigen Ergebnissen erschwert bzw. eingeschränkt wird.

So wird man bei der Suche auf healthline.com bei jedem Besuch darauf hingewiesen, dass nun in den ‚weltbesten‘ Gesundheitsinformationen gesucht wird, wie diese unbescheidene Einschätzung zustande kommt, wird aber nicht weiter erwähnt. Wer definiert, welche Gesundheitsseite zu den weltbesten gehört?

Das oft geforderte Kriterium der Unabhängigkeit ist für vertikale Suchdienste wesentlich schwieriger zu erreichen. Trotzdem sind vertikale Suchdienste ein Weg, um die Qualität zu verbessern, vorausgesetzt es wird sorgsam vorgegangen und Bedacht darauf genommen, dass das kommerzielle Interesse nicht das einzige Interesse ist.

Ein auffälliger Unterschied besteht in der Beurteilung der Zufriedenheit mit dem, was das Internet an Informationen über Gesundheit bereithält. Wie in dieser Studie mehrfach erwähnt, sind heute die Nutzer, Health-Seeker, zum überwiegenden Teil zufrieden mit dem, was sie finden.

Im Gegensatz dazu sagen Entscheidungsträger (N=53) von US-basierten Gesundheitsdiensten in einer 2005 durchgeführten Befragung im Einklang, dass die Qualitätskontrolle verbessert werden müsse. Laut den Autoren dieser Studie erscheint den Befragten die Notwendigkeit zur Entwicklung von Tools zur besseren Kenntlichmachung bzw. Regulierung Gesundheitswebseiten hohe Priorität zu besitzen¹¹⁶.

Trotzdem ist die Tatsache, dass die Zufriedenheit bei den Health-Seekern der Gesundheitsinformationen selbst in den letzten Jahren zugenommen hat, auffällig.

Internetbenutzer haben heute wesentlich höheres Vertrauen in die Gesundheitsinformationen die im Web dargeboten werden als noch vor einigen Jahren¹¹⁷.

1999 fürchteten sich 86% der Amerikaner vor Gesundheitsinformationen, die aus unzuverlässigen Quellen stammen. 2006 sagen die meisten Amerikaner, sie sind zufrieden mit dem was sie finden, nur wenige finden die Informationen verwirrend oder frustrierend. 1999 suchten noch 56% aller Nutzer nach Zusatzinformationen, um

¹¹⁶ Vgl. Lorence, D. P. / L. Greenberg, 'The zeitgeist of online health search. Implications for a consumer-centric health system' (2006) 21(2) *J Gen Intern Med* 134-139, S. 136-137.

¹¹⁷ Vgl. Fox, Susanna *Online Health Search 2006* (2006)

<http://www.pewinternet.org/pdfs/PIP_Online_Health_2006.pdf> am 03.03.2007 , S. 9.

festzustellen, wie glaubwürdig die gefundenen Informationen sind¹¹⁸. 2006 sinkt der Anteil von Nutzern die immer oder fast immer die Herkunft und das Datum der Information prüfen auf 25%.¹¹⁹

Die in den USA festgestellte Erosion der Aufmerksamkeit der Health-Seeker gegenüber den Qualitätsindikatoren wie Herkunft und Aktualität¹²⁰ hat einen handfesten Grund. Wie die Studie von Susanna Fox nachweist, stellt nur ein sehr geringer Anteil an Webdiensten solche Informationen überhaupt dar.¹²¹

4.2 Wikipedia als Ressource für Gesundheitsinformationen

Interessanterweise wurde Wikipedia von der einschlägigen Wissenschaft bis jetzt nicht als Quelle für Gesundheitsinformationen erkannt, obwohl solche Informationen, in der deutschsprachigen Version der Online-Enzyklopädie, aber mit überaus hoher Wahrscheinlichkeit auch in der englischen Version, bereits jahrelang vorhanden sind.

Das mag damit zusammenhängen, dass eine Enzyklopädie aufgrund ihres lexikalischen Charakters nicht als umfassende Informationsquelle für die Abfrage von Gesundheitsinformationen durch Laien eingestuft wird.

Die Orthopäden Mathur et al etwa verwendeten eine ähnliche Methode, wie sie auch bei der Feldanalyse in dieser Studie verwendet worden ist. Sie fragten im Jahr 2005 den Suchbegriff *scoliosis* bei 5 Suchdiensten ab und ermittelten die Qualität von 50 Websites zum Thema Skoliose, die die Suchmaschinen zurücklieferten. Die allgemeine Qualität der Information über Skoliose wurde als limitiert bezeichnet. Es wurde hervorgehoben, dass sowohl private Websites von Selbsthilfegruppen als auch Websites mit medizinisch-akademischen Hintergrund deutliche Qualitätsmängel aufwiesen.¹²²

Interessanterweise wurde Wikipedia in dieser Studie nicht als entsprechende Quelle für Informationen über Skoliose genannt, obwohl mit großer Wahrscheinlichkeit davon

¹¹⁸ Vgl. Fox, Susanna / Lee Rainie, *The Online Healthcare Revolution: How the Web helps Americans take better care of themselves* (1999) <http://www.pewinternet.org/pdfs/PIP_Health_Report.pdf> am 03.03.2007, S. 14.

¹¹⁹ Vgl. Fox, Susanna *Online Health Search 2006* (2006) <http://www.pewinternet.org/pdfs/PIP_Online_Health_2006.pdf> am 03.03.2007, S. 11.

¹²⁰ Vgl. Kap. 3.1.3, S. 32.

¹²¹ Vgl. Fox, Susanna *Online Health Search 2006* (2006) <http://www.pewinternet.org/pdfs/PIP_Online_Health_2006.pdf> am 03.03.2007, S.11-12.

¹²² Vgl. Mathur, S. et al, 'Surfing for scoliosis: the quality of information available on the Internet' (2005) 30(23) *Spine* 2695-2700.

ausgegangen werden kann, dass die URL des Skoliose-Artikels in den Suchmaschinen angezeigt wurde, existiert der Artikel in der englischsprachigen Wikipedia doch seit 23. Oktober 2002.

Der Umfang der Wikipedia geht mittlerweile über den einer klassischen Enzyklopädie weit hinaus. Der Großteil der untersuchten medizinischen Artikel ist beispielsweise sehr umfassend und auch allgemein verständlich abgefasst.

So ist eine der wenigen prinzipiellen Richtlinien für Autoren der Wikipedia die Ausgewogenheit der Information. Diese Ausgewogenheit ist es wohl auch, die gerade für Darstellung von medizinischen Informationen eine hohe Glaubwürdigkeit besitzt.

Selbstheilungskraft oder die evolutive Artikelentwicklung durch die Community ist die grundlegendste Eigenschaft von Wikipedia. Jeder Autor trägt sein Wissen bei und erzeugt so mehr, als von der Summe der Einzelinformationen zu erwarten wäre.

Dass laut Ergebnis dieser Analyse gesundheitsrelevante Artikel in Wikipedia überwiegend als erstgereichte Einträge bei der größten Suchmaschine gefunden werden, darf man tatsächlich von einer wichtigen und oft genutzten Informationsquelle ausgehen.

Auch ist es im Umkehrschluss nahe liegend, dass gesundheitsrelevante Artikel in der englischsprachigen Wikipedia ebenso prominente Plätze in den großen Suchmaschinen einnehmen wie in der deutschsprachigen Wikipedia.

Stellt man darüber hinaus die durchschnittlich Seitenbearbeitungszahlen der deutschen Wikipedia von 20,91 mit den Seitenbearbeitungszahlen der 60 untersuchten Gesundheits-Artikel gegenüber, beobachtet man, dass diese um den Faktor 8 bis 18¹²³ höher liegen. Die Bearbeitungszahlen der Gesundheitsartikel liegen also weit über der durchschnittlichen Seitenbearbeitungszahl aller Wikipedia-Artikel.

Ursprünglich war der Aufbau der in dieser Arbeit durchgeführten Feldstudie so geplant, dass medizinische Artikel in der Wikipedia selbst angelegt werden sollten, um die evolutive Entwicklung zu verfolgen und zu dokumentieren.

Es stellte sich aber während der ersten Prüfung der Durchführbarkeit rasch heraus, dass zu praktisch allen Krankheitsbegriffen bereits Artikel in der Wikipedia vorhanden waren.

¹²³ Vgl. Kap. 3.2.2 S. 35.

Trustability liegt in der Natur des Wikipedia selbst, da es als eine der 100 bekanntesten Websites, derzeit liegt sie auf Platz 10¹²⁴, und als NGO in allen Gesellschaftsschichten und Kulturen anerkannt ist. Da Wikipedia in vielen Sprachen verbreitet ist, dürfte es hier auch besonders leicht sein, Informationen in Regionen zu bringen, die bisher unterversorgt waren.

Gerade solche Informationen sind auch für Volkswirtschaften erstrebenswert, kann doch durch richtiges Verhalten - beispielsweise bei Masseninfektionen - großes Verhinderungspotenzial erzielt und so großer Schaden von der Bevölkerung abgehalten werden.

Eine der wichtigsten von Qualitätsinitiativen geforderten Kriterien für das Bereitstellen von Gesundheitsinformationen, die Nennung der Autoren der betreffenden Information, kann Wikipedia nur in anonymisierter Form bieten, denn das kollaborative Medium setzt auf die Gestaltungskraft der Masse.

Bestand hat, was von der Community akzeptiert wird. Im Gegensatz zum von Wikipedia gewählten Ansatz tragen immer noch wenige Websites, die Gesundheitsinformationen publizieren irgendwelche Quellenangaben, Erstellungsdaten oder andere Qualitätsindikatoren.¹²⁵ Der Ansatz von Wikipedia scheint also auch hier der effektivere zu sein.

Die von HON unter dem Kriterium *Transparency* geforderte Erreichbarkeit des Autors via E-Mail stammt aus der Frühzeit des Internet und kann heute auch mit anderen Mitteln, etwa formularbasiert im World Wide Web, erreicht werden. So ist es in Wikipedia besonders leicht über die Diskussionsseiten des jeweiligen Artikels mit jedem publizierenden Autor in Kontakt zu treten.

Offensichtlich erfüllt ein kollaboratives System wie Wikipedia praktisch alle Anforderungen für die glaubwürdige Darstellung von Gesundheitsinformationen im Web. Das hat diese Analyse in der deutschsprachigen Wikipedia eindeutig ergeben.

¹²⁴ Vgl. o.V., *Alexa Top 500 Web Sites* (2007) Alexa Internet, Inc.

<http://www.alexa.com/data/details/traffic_details?url=wikipedia.org> am 10.05.2007 .

¹²⁵ Vgl. Fox, *Susanna Online Health Search 2006* (2006)

<http://www.pewinternet.org/pdfs/PIP_Online_Health_2006.pdf> am 03.03.2007 , S. 12.

Die medizinische Analyse in dieser Arbeit legt den Schluss nahe, dass praktisch an allen Artikeln auch Mediziner mitgewirkt haben müssen, ja manche Artikel sogar in Stile eines Medizinlehrbuchs aufgebaut sind.

Nach diesen Erkenntnissen und aktuellen Informationen über die deutsche Wikipedia, dürfte die Qualität der Artikel vor allem im naturwissenschaftlichen Bereich tatsächlich auf das verstärkte Engagement von Wissenschaftlern zurückzuführen sein.¹²⁶

Hier ist vielleicht auch die Schwäche zu erkennen, dass nicht alle Artikel gleichermaßen auch für Laien geeignet sind. Spezialisierte Webdienste von Selbsthilfegruppen oder auch von kommerziellen Anbietern könnten hier eventuell besser zugeschnittene Information bieten. Ob und in welchem Maß bereits jetzt Mitglieder von Selbsthilfegruppen am Kollaborationskollektiv teilnehmen, konnte in dieser Arbeit nicht geklärt werden. Weiterführende Untersuchungen sind wünschenswert.

Tatsache ist, dass die Entwicklung von staatlich geförderten Qualitätsinitiativen nicht die Breitenwirksamkeit erreicht hat, die man sich seitens der Entwickler und Unterstützer erwartet hatte.

Ob es Zufall ist, oder ob es mehr als einen zeitlichen Zusammenhang zwischen der Verlangsamung der Entwicklung dieser Initiativen und dem Aufkommen von Gesundheitsinformationen aus kollaborativen Wissenssystemen gibt, kann nicht geklärt werden.

Möglicherweise wurde durch das Platzen der Dotcom-Blase und der damit einhergehenden Rückbesinnung auf die Grundwerte des World Wide Web die Entwicklung frei zugänglicher kollaborativer Wissenssysteme unterstützt. Das Auftreten von Systemen, die stark auf die Mitwirkung der Community bauen, setzte jedenfalls zeitnah ein.

Offensichtlich ist, dass Wikipedia heute die größte verfügbare Wissenssammlung mit frei zugänglicher Information ist. Eine Förderung dieser auf finanzielle Unterstützung angewiesenen Non-Profit-Organisation ist also im höchsten Maß zu empfehlen.

4.3 Schlussfolgerungen der medizinischen Analyse

Die Ergebnisse der Feldanalyse scheinen eine oft angeführte Behauptung, allgemeine Suchmaschinen liefern eben so breite, unspezifische Ergebnisse zurück, nicht zu

¹²⁶ Vgl. Kleinz, Torsten, *Kritik und Unterstützung für Wikipedia* (2007)

<<http://www.heise.de/newsticker/meldung/88359>> am 23.04.2007 .

bestätigen, zumindest nicht für die Suche nach medizinischen Krankheitsbegriffen in dieser Arbeit.

Sim et al haben nachgewiesen, dass typischerweise die einfachen Krankheitsbegriffe, wie Symptome, Zustand (Name der Krankheit) von den Health-Seekern zur Abfrage eingegeben werden¹²⁷.

Eben diese Eingabe von einfachen und verbreiteten Begriffen lieferte auch in dieser Analyse hochrelevante Ergebnisse wieder, sodass die große Akzeptanz der User gegenüber den Suchmaschinen als Einstiegspunkt für die Suche nach Gesundheitsinformationen einen berechtigten Grund hat.

4.4 Soziale Auswirkungen

Es gibt viele Anzeichen, dass die kollektive Intelligenz das Web und im zunehmendem Maße auch die Gesellschaft nachhaltig verändern wird.

Die Änderung der Gesellschaft, die von Philosophen und Sozialforschern als Transformation der Industriegesellschaft in eine Wissens- und Informationsgesellschaft postuliert wurde, ist zum Teil bereits Realität geworden.

Heute gilt es vermehrt Anstrengungen zu unternehmen, für alle sozialen Gruppen der Weltbevölkerung einen Zugang zu Online-Informationen zu schaffen, um das derzeitige Ungleichgewicht aufzuheben. Hier gilt es nicht nur, den Zugang zu den neuen Informationsquellen herzustellen, es muss auch sichergestellt werden, dass die Menschen auch befähigt werden, mit den neuen Techniken und Informationen umzugehen.

4.5 Ausblick auf Semantic Web Anwendungen

Semantic Web Anwendungen sind 2006 in einer noch überschaubaren Anzahl vorhanden. Das von Berners-Lee postulierte Abheben findet statt, wenn auch etwas später als prognostiziert.

Die Ursache für die Verspätung könnte wiederum in der lockeren Spezifikation des W3C liegen, die den Entwicklern und Ingenieuren viel Freiraum lässt und die dem Grundprinzip von Tim Berners-Lee folgt, so wenig Regeln wie möglich aufzustellen.¹²⁸

¹²⁷ Vgl. Sim, Nurain Z. et al, 'Information on the World Wide Web—how useful is it for parents?' (2007) 42 *Journal of Pediatric Surgery* 305-312, S. 306.

¹²⁸ Vgl. Berners-Lee, Tim / Mark Fischetti, *Weaving the Web : The Past, Present and Future of the World Wide Web by its inventor* (1999).

Eine mögliche weitere Erklärung für die Verzögerung liefert auch Dan Brickley mit seiner Neuinterpretation des Semantic Web Layer¹²⁹ (Abbildung 17).

Brickley arbeitete bis 2005 für das W3C, vor allem für die Standardisierung des RDF. Im Jahr 2000 arbeitete er auch mit Eysenbach bei der Entwicklung des MedCERTAIN Projektes¹³⁰ zusammen.

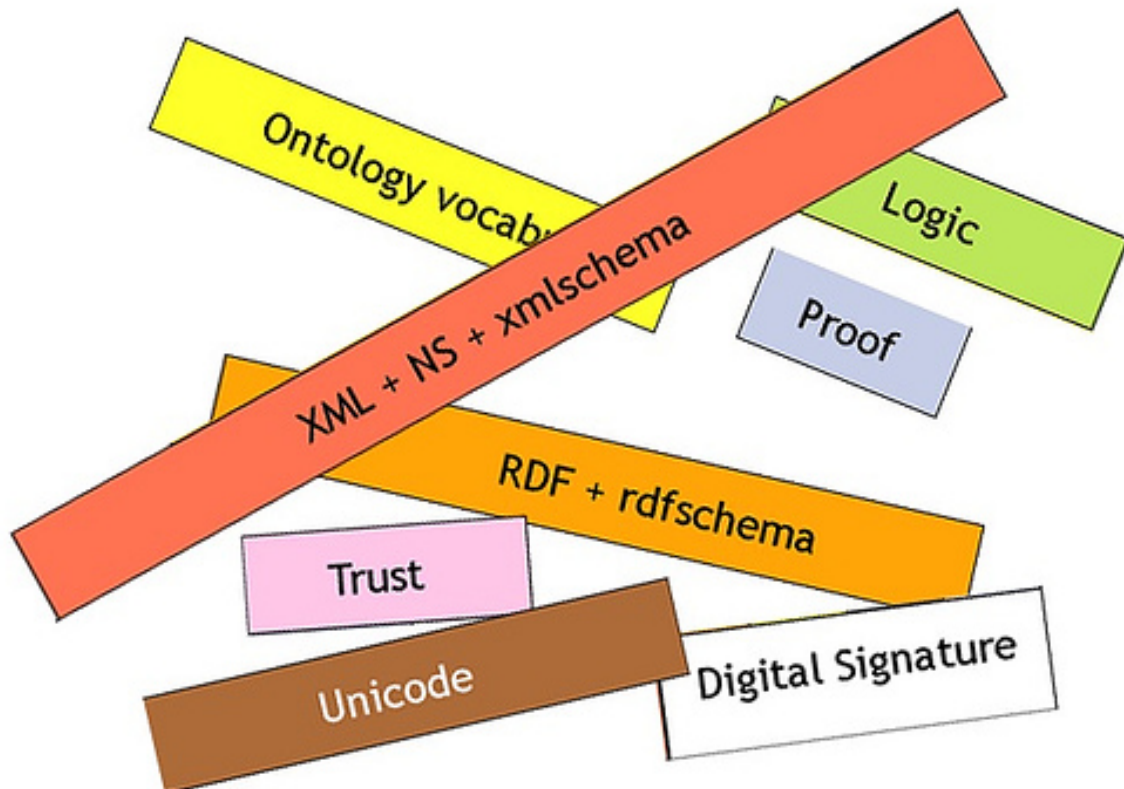


Abbildung 17: The truth about the Semantic Web (nach Brickley, 2007)

Auf breiter Ebene haben sich Semantic Web Technologien bis jetzt noch nicht durchgesetzt. Grigoris Antoniou und Frank van Harmelen sagten 2004 dazu: „*Still, the question remains why users should care, why they should abandon HTML for Semantic Web Languages*“¹³¹.

¹²⁹ Vgl. Brickley, Dan, *The truth about the semantic Web...* (2007)

<<http://www.flickr.com/photos/danbri/428172848>> am 22.04.2007 .

¹³⁰ Vgl. Kap. 3.4.3, S.48.

¹³¹ Antoniou, G. / Frank van Harmelen, *A semantic Web primer*, Cooperative information systems (2004),S. 9.

Eine weitere Meinung bringt Lara Slaughter 2006 zum Ausdruck. Obwohl klar ist, dass Semantic Information die Leistungsfähigkeit von Abfragesystemen verbessern kann, gibt es keinen Beleg dafür, dass Semantic Web Technologien dem Health-Seeker helfen könnten, medizinische Informationen besser zu verstehen oder dadurch gezielter Fragen stellen zu können¹³².

Aber wie HTML von immer mehr und mehr Entwicklern akzeptiert worden ist und zum Standard wurde, so hat sich aus der Familie der Semantic Web Werkzeuge zumindest XML in den letzten Jahren stark etabliert. XML für sich gesehen kann sicher nicht als die Realisation des Semantic Web angesehen werden, aber es ist ein wichtiger Schritt in diese Richtung.

Die aktuell größte Herausforderung ist nicht die Annahme des Themas durch die Wissenschaft, sondern die Umsetzung durch Technologieentwickler und Softwareingenieure.¹³³

Einer breiten Öffentlichkeit ist der Begriff Semantic Web eher schwer vermittelbar, darum findet man andernorts gelegentlich die Bezeichnung Web 3.0 dafür. Dieser an das Web 2.0 angelegte Begriff ist aber weder definiert noch durch eine weit verbreitete Verwendung legitimiert.

4.6 Die Zukunft der klassischen Qualitätsinitiativen

Die etablierten Initiativen zur Sicherung der Qualität von Gesundheitsinformationen werden aus heutiger Sicht eine Randerscheinung bleiben.

Die Regulierung der Inhalte des World Wide Web ist aufgrund der Größe und der Dynamik praktisch nicht durchführbar. Sie ist auch nicht erstrebenswert.

Zitat Purcell:

“Regulation does not seem the right strategy for improving the quality of health information on the internet. Other approaches, such as educating the producers of this content, look like a better long term bet. However, such initiatives should not hinder

¹³² Vgl. Slaughter, L. A. / D. Soergel / T. C. Rindflesch, 'Semantic representation of consumer questions and physician answers' (2006) 75(7) *Int J Med Inform* 513-529, S. 514.

¹³³ Vgl. Antoniou, G. / Frank van Harmelen, *A semantic Web primer*, Cooperative information systems (2004), S. 9.

the evolution of communities, resources and processes that are improving healthcare outcomes.”¹³⁴

Initiativen mit dem Anspruch zur Regulierung sind bis auf die Systeme mit Gütesiegelvergabe nicht mehr vorhanden. In der einschlägigen Literatur hat es bis zur Jahrtausendwende zahlreiche Studien und Experimente zur Regulierung von Gesundheitsinformationen gegeben, in den letzten Jahren ist die Zahl der Arbeiten aber deutlich zurückgegangen.

Ein weiterer Grund für das Stagnieren dürfte auch die Tatsache sein, dass Benutzer heute ein höheres Vertrauen in die dargebotenen Gesundheitsinformationen haben, als dies noch vor einigen Jahren der Fall war.

Gütesiegelsysteme sind aus Sicht des Benutzers einfach und effektiv, können aber zur Abstumpfung der Kritikfähigkeit gegenüber einem Informationsanbieter führen. Solche Systeme waren zu Beginn nicht fälschungssicher, da Logos oft ohne Autorisierung kopiert und verwendet wurden. Mittlerweile ist die Technologie hinter dem Gütesiegel aber wesentlich weiter ausgereift. Über das Logo gelangt man zu den Daten der Zertifizierungsstelle, die den Status des Zertifikates an den Benutzer zurückmelden.

4.7 Limitationen

Die Ergebnisse dieser Arbeit sind einigen Einschränkungen unterworfen. Studien von US Autoren wurden auf europäische Verhältnisse umgelegt und die deutsche Wikipedia wurde als repräsentativ für die englischsprachige Wikipedia angenommen. Obwohl für beide Annahmen gute Gründe sprechen, sind gewisse Unschärfen damit nicht ausgeschlossen.

Naturgemäß konnte das Thema nicht in voller Breite und Tiefe beleuchtet werden. Zahlreiche Arbeiten, die mir zur Verfügung standen, konnten in dem geforderten Rahmen nicht ausreichend berücksichtigt werden. Sicher existieren noch Studien zu diesem Thema, die ich während meiner Recherche nicht entdeckt habe.

Auch die Feldanalyse musste methodisch eingeschränkt werden. So wurde etwa die Abfrage der Krankheitsbegriffe auf den Suchdienst Google beschränkt, zweifellos der derzeit wichtigste, aber nicht einzige Suchdienst.

¹³⁴ Purcell, G. P. / P. Wilson / T. Delamothe, 'The quality of health information on the internet' (2002) 324(7337) *Bmj* 557-558, S. 557.

Eine Einschränkung darauf scheint aber gerechtfertigt, da alle ausgewerteten Studien, europäische wie amerikanische, einen Anteil von mehr als 70% für diesen Suchdienst bei der Abfrage von Gesundheitsinformationen ausweisen.

Es war mir ein Anliegen - neben dem Nachweis, den die Feldanalyse erbracht hat - das Thema in einem größeren Zusammenhang darzustellen und technologische Entwicklungen, zukünftige wie vergangene, anzusprechen.

Begriffsdefinitionen sollten dem Leser helfen, das Thema so zu erfassen, wie ich es beabsichtigt habe. Dabei kann ich keinen Anspruch auf Vollständigkeit der Definitionen erheben.

5 Conclusio

Viele Studien sowie die Ergebnisse dieser Arbeit zeigen, dass die Nachfrage nach Gesundheitsinformationen im World Wide Web ungebrochen und sehr hoch ist.

Der übliche Weg, den ein Health-Seeker nimmt, um zu Gesundheitsinformationen zu gelangen, ist der Weg über einen der großen Suchdienste.

Die Information dient in zunehmendem Maß nicht nur den Suchenden selbst, vermehrt wird auch im Auftrag oder zum Wohl der Angehörigen gesucht. Die üblichen Suchbegriffe sind einfache und kurze Abfragestrings, die entweder die Krankheit bzw. den Zustand des Erkrankten bezeichnen. Wichtig sind den Benutzern nicht nur Informationen zum gesund werden, sondern auch solche zum gesund bleiben.

Das Vertrauen in die dargebotenen Gesundheitsinformationen im Web ist heute höher als vor einigen Jahren. Gleichzeitig erodiert die Aufmerksamkeit, speziell im amerikanischen Raum, gegenüber offensichtlichen und leicht festzustellenden Qualitätskriterien wie Autorenschaft und Datum der Aktualisierung.

5.1 Überprüfung der zentralen Fragestellungen

Die Ergebnisse dieser Studie zeigen, dass die von Fachleuten propagierten Technologien, um die Qualität solcher Gesundheitsinformationen zu erhöhen bzw. zu regulieren, nicht in dem Maß verbreitet sind, dass dadurch bereits eine nachhaltige Verbesserung des Informationsgehalts erzielt werden würde.

Applikationen, die heute hauptsächlich mit Semantic Web Technologie entwickelt werden, haben den Experimentalstatus oft noch nicht verlassen.

Kollaborative Wissenssysteme wie Wikipedia sind weit verbreitet und akzeptiert. Von Beginn an wurde diese Online-Enzyklopädie auch benutzt, um Gesundheitsinformationen darzustellen. Solche spezifischen Artikel aus der Wikipedia nehmen in den Ergebnislisten der Suchdienste prominente Platzierungen ein. Von hohen Zugriffsraten und einer intensiven Nutzung dieser Informationsquelle ist daher auszugehen.

Die untersuchte Qualität der medizinischen Information in der deutschsprachigen Wikipedia zeigte durchwegs gute bis sehr gute Qualität hinsichtlich Umfang, Richtigkeit sowie Verständlichkeit für Laien.

Wikipedia scheint also für die Benutzer eine wichtige und ernstzunehmende Ressource für die Abfrage von Gesundheitsinformation zu sein. Interessanterweise hat sich noch keine Studie mit diesem Aspekt beschäftigt.

Die Nachfrage nach Gesundheitsinformationen ist in den Ländern mit hoher Internet-Durchdringung steigend.

Da das Internet vor allem in den Schwellenländern zunehmend an Bedeutung gewinnt, ist davon auszugehen, dass auch dort die Nachfrage nach Gesundheitsinformationen stark an Bedeutung zunehmen wird.

Da Wikipedia in mehr als 100 Sprachen verfügbar ist, ist es in diesem Medium besonders einfach, Informationen auch in solche Ländern zu bringen, die bisher aufgrund von Sprachbarrieren kaum relevante Informationen aus dem Internet beziehen konnten.

5.2 Ausblick

Das World Wide Web wird bleiben. Weder das Web 2.0 noch das Semantic Web werden das World Wide Web ersetzen oder parallel dazu existieren. Technologien und Nutzung entwickeln sich ständig weiter, aber diese Entwicklungen werden innerhalb des World Wide Web stattfinden. Nur die Idee des World Wide Web garantiert einen freien Zugang zur Information. Information und Wissen werden maßgeblich die Zukunft unserer Gesellschaft beeinflussen.

Das Semantic Web wird die Zukunft des World Wide Web verändern, die Verfügbarkeit, und die Glaubwürdigkeit von Information und Wissen kann damit erhöht werden. Das kollaborative Web hat seinen Platz erobert und bereits einen wesentlichen Beitrag geleistet um glaubwürdige, unverfälschte Informationen zu verbreiten.

Regulierungsversuche und Filtertechnologien im World Wide Web haben nicht zur Verbesserung der Qualität von Gesundheitsinformationen beigetragen. Regulierung sollte auf das notwendige Maß - wie das W3C es vorlebt - beschränkt sein, denn wir wollen nicht, dass das Zeitalter der Informationsgesellschaft einhergeht mit einer weiteren Beschränkung der Bürgerrechte.

Der Kampf gegen die übermäßige Regulierung des World Wide Webs wird noch zu führen sein. Es ist nicht anzunehmen, dass wichtige Entscheidungsträger immer so agieren, wie es Berners-Lee propagiert.

Nutzen wir die Chance, die sich uns bietet, überwinden wir soziale Barrieren und grenzen wir zum Wohle der Vielfalt niemanden vom Zugang zu Wissenssystemen aus.

5.3 Empfehlungen

Das Projekt Wikipedia hat einen sehr bemerkenswerten Verlauf genommen. In exponentiellen Schritten wurde und wird Wissen und Information aus allen Bereichen des menschlichen Daseins allen Online-Benutzern kostenlos und frei zur Verfügung gestellt. Der Betreiber agiert ohne Gewinnabsicht. Die Wikipedia Foundation ist daher auf unser aller Mithilfe angewiesen.

Der Einzelne profitiert von Wissen, das dort durch andere verfügbar ist, der Staat profitiert durch informierte Bürger. Dieses Projekt kann aber nur dann weiter existieren, wenn die Kosten dafür aufgebracht werden. Hier ist die öffentliche Hand in der Pflicht, diese Art der Wissenssysteme finanziell zu unterstützen.

Zusammenfassend sind Maßnahmen und weiterführende Forschungsaktivitäten für folgende Bereiche zu empfehlen:

- Sponsoring von kollaborativen Systemen durch die zuständigen Institutionen
- Einbeziehung von Semantic Web Technologie in aktuellen Entwicklungen
- Monitoring und wissenschaftliche Betrachtung des Kollaborationskollektivs
- Gezielte Befragungen der Bevölkerung bezüglich des Verhaltens bei der Suche nach Gesundheitsinformationen im World Wide Web
- Entwicklung von *Shared Decision Making* Modellen im Gesundheitswesen
- Berücksichtigung des Nutzungsverhaltens der Bürger bei der Entwicklung von Gesundheitsportalen
- Monitoring und Überblick über vertikale Suchmaschinen, die auf Gesundheitsinformationen spezialisiert sind
- Vergleich der Gesundheitsinformationen in der deutschen und der englischen Wikipedia

6 Anhang

Der Datenauszug der Online Abfrage der Datenbank von Eurostat ist im diesem Anhang in tabellarischer Form mit Zugriffsdatum und Aktualisierungsinformationen wiedergegeben.

6.1 Anteil der Internet-Zugänge von Haushalten der EU (%)

Datenabfrage eurostat_2006_internet in Haushalten der EU 25

Datum des Auszugs: Sun, 11 Mar 07 11:09:36

Letzte Aktualisierung: Thu Feb 22 15:05:16 MET 2007

Geo	2002	2003	2004	2005	2006
eu15 Europäische Union (15 Länder)	39	43	45	53	54
be Belgien	:	:	:	50	54
bg Bulgarien	:	:	10	:	17
cz Tschechische Republik	:	15	19	19	29
dk Dänemark	56	64	69	75	79
de Deutschland	46	54	60	62	67
ee Estland	:	:	31	39	46
ie Irland	:	36	40	47	50
gr Griechenland	12	16	17	22	23
es Spanien	:	28	34	36	39
fr Frankreich	23	31	34	:	41
it Italien	34	32	34	39	40
cy Zypern	24	29	53	32	37
lv Lettland	3	:	15	31	42
lt Litauen	4	6	12	16	35
lu Luxemburg (Grand-Duché)	40	45	59	65	70
hu Ungarn	:	:	14	22	32
mt Malta	:	:	:	:	53
nl Niederlande	58	61	:	78	80
at Österreich	33	37	45	47	52
pl Polen	11	14	26	30	36
pt Portugal	15	22	26	31	35
ro Rumänien	:	:	6	:	:
si Slowenien	:	:	47	48	54
sk Slowakei	:	:	23	23	27
fi Finnland	44	47	51	54	65
se Schweden	:	:	:	73	77
uk Vereinigtes Königreich	50	55	56	60	63

Tabelle 8: Anteil der Internetzugänge in EU-Haushalten (Eurostat)

6.2 Internetaktivitäten - Einzelpersonen

Datum des Auszugs: Thu, 3 May 07 03:13:53

Letzte Aktualisierung: Tue Apr 10 14:15:15 MEST 2007

6.2.1 Prozent der Einzelpersonen, Jahr 2006

Quelle: Prozent der Einzelpersonen

Prozentualer Anteil der Personen, die das Internet in den letzten 3 Monaten zur Kommunikation genutzt haben:

	EU25	DE	AT	UK	NO
Alle Personen	46	62	54	54	73
Personen, 16 bis 24 Jahre	73	89	78	69	94
Personen, 25 bis 34 Jahre	61	83	71	67	89
Personen, 35 bis 44 Jahre	52	72	61	63	80
Personen, 45 bis 54 Jahre	54	73	62	66	81
Personen, 55 bis 64 Jahre	42	61	52	55	70
Personen, 65 bis 74 Jahre	28	43	32	39	59
Personen, 75 bis 84 Jahre	11	20	12	20	23

Tabelle 9: Anteil der Einzelpersonen in den letzten drei Monaten (Eurostat)

6.2.2 Prozent der Internetnutzer, 2006

Quelle: Prozent der Einzelpersonen die in den letzten drei Monaten das Internet benutzen haben.

Prozentualer Anteil der Personen, die das Internet in den letzten 3 Monaten zur Kommunikation genutzt haben.

	EU25	DE	AT	UK	NO
Alle Personen	85	90	88	82	90
Personen, 16 bis 24 Jahre	88	94	90	83	95
Personen, 25 bis 34 Jahre	87	93	89	84	93
Personen, 35 bis 44 Jahre	85	90	88	84	89
Personen, 45 bis 54 Jahre	84	89	87	86	90
Personen, 55 bis 64 Jahre	82	87	86	81	84
Personen, 65 bis 74 Jahre	81	86	87	75	89
Personen, 75 bis 84 Jahre	83	84	86	83	80

Tabelle 10: Anteil der Internetnutzer in den letzten drei Monaten (Eurostat)

6.3 Gesundheitsrelevante Internetaktivitäten, 2006

Datum des Auszugs: Thu, 3 May 07 03:13:53

Letzte Aktualisierung: Tue Apr 10 14:15:15 MEST 2007

6.3.1 Prozent der Einzelpersonen, 2006

Quelle: Prozent der Einzelpersonen

Prozentualer Anteil der Personen, die das Internet für die Beschaffung von gesundheitsrelevanten Informationen über Verletzungen, Krankheiten oder Ernährung genutzt haben:

		EU25	DE	AT	UK	NO
Alle Personen		46	62	54	54	73
Personen, 16 bis 24 Jahre		73	89	78	69	94
Personen, 25 bis 34 Jahre		61	83	71	67	89
Personen, 25 bis 54 Jahre		52	72	61	63	80
Personen, 35 bis 44 Jahre		54	73	62	66	81
Personen, 45 bis 54 Jahre		42	61	52	55	70
Personen, 55 bis 64 Jahre		28	43	32	39	59
Personen, 65 bis 74 Jahre		11	20	12	20	23

Tabelle 11: Anteil der der Health-Seeker (Eurostat)

6.3.2 Prozent der Internetnutzer, 2006

Quelle: Prozent der Einzelpersonen die in den letzten drei Monaten das Internet benutzen haben

Prozentualer Anteil der Personen, die das Internet für die Beschaffung von gesundheitsrelevanten Informationen über Verletzungen, Krankheiten oder Ernährung genutzt haben:

		EU25	DE	AT	UK	NO
Alle Personen		85	90	88	82	90
Personen, 16 bis 24 Jahre		88	94	90	83	95
Personen, 25 bis 34 Jahre		87	93	89	84	93
Personen, 25 bis 54 Jahre		85	90	88	84	89
Personen, 35 bis 44 Jahre		84	89	87	86	90
Personen, 45 bis 54 Jahre		82	87	86	81	84
Personen, 55 bis 64 Jahre		81	86	87	75	89
Personen, 65 bis 74 Jahre		83	84	86	83	80

Tabelle 12: Anteil der Health-Seeker in den letzten drei Monaten (Eurostat)

7 Literaturverzeichnis

7.1 Fachliteratur:

Andreassen, H. K. et al, 'Health-related use of the Internet in the Norwegian population' (2006) 126(22) *Tidsskr Nor Laegeforen* 2950-2952

Antoniou, G. und van Harmelen, Frank, *A semantic Web primer*, Cooperative information systems (2004)

Ayantunde, A. A., Welch, N. T. und Parsons, S. L., 'A survey of patient satisfaction and use of the Internet for health information' (2007) 61(3) *Int J Clin Pract* 458-462

Berners-Lee, Tim und Fischetti, Mark, *Weaving the Web : The Past, Present and Future of the World Wide Web by its inventor* (1999)

Boulos, M. N. Kamel , Maramba, I. und Wheeler, S., 'Wikis, blogs and podcasts: a new generation of Web-based tools for virtual collaborative clinical practice and education' (2006) 6 *BMC Med Educ* 41

Boyer, C. und Geissbuhler, A., 'A decade devoted to improving online health information quality' (2005) 116 *Stud Health Technol Inform* 891-896

Brin, Sergey und Page, Lawrence, 'The anatomy of a large-scale hypertextual {Web} search engine' (1998) 30 *Computer Networks and ISDN Systems* 107-117

Eysenbach, Gunther Diepgen, Thomas L, 'Towards quality management of medical information on the internet: evaluation, labelling, and filtering of information' (1998) 317 *Bmj* 1496-1502

Eysenbach, Gunther et al, 'MedCertain: Quality Management, Certification and Rating of Health Information on the Net' (Arbeit präsentiert bei den AMIA 2000 Annual Symposion, Santa Anita, A, 2000)

Giles, Jim, 'Internet encyclopaedias go head to head' (2005) 438 *Nature* 900-901

Hesse, Bradford W et al, 'Trust and Sources of Health Information. The Impact of the Internet and Its Implications for Health Care Providers: Findings From the First Health Information National Trends Survey' (2005) 165(22) *Arch Intern Med* 2618-2624

Jadad, Alejandro R Gagliardi, Anna, 'Rating Health Information on the Internet - Navigating to Knowledge or to Babel?' (1998) 279 *Jama* 611 - 614

Kaplan, Craig A., 'Collective Intelligence: a new approach to stock price forecasting' (Arbeit präsentiert bei den Proceedings of the 2001 IEEE Systems, Man, and Cybernetics Conference, Tucson, AZ, USA, 2001)

Koivunen, Marja-Riitta und Miller, Eric, 'W3C Semantic Web Activity' (Arbeit präsentiert bei den Semantic Web Kick-off Seminar in Finland, Helsinki, 2001)

Lorenz, D. P. und Greenberg, L., 'The zeitgeist of online health search. Implications for a consumer-centric health system' (2006) 21(2) *J Gen Intern Med* 134-139

Mathur, S. et al, 'Surfing for scoliosis: the quality of information available on the Internet' (2005) 30(23) *Spine* 2695-2700

Mayer, M. A. et al, 'MedCIRCLE: collaboration for Internet rating, certification, labelling and evaluation of health information on the World-Wide-Web' (2003) 95 *Stud Health Technol Inform* 667-672

McLellan, Faith, 'Like hunger, like thirst: patients, journals, and the internet' (1998) Volume 352, Supplement 2 *The Lancet* S39-S43

Musser, John, *Web 2.0 Principles and Best Practices* (2006)

Purcell, G. P., Wilson, P. und Delamothé, T., 'The quality of health information on the internet' (2002) 324(7337) *Bmj* 557-558

Risk, Ahmad und Dzenowagis, Joan, 'Review of internet health information quality initiatives' (2001) 3(4) *J Med Internet Res* E28

Sim, Nurain Z. et al, 'Information on the World Wide Web—how useful is it for parents?' (2007) 42 *Journal of Pediatric Surgery* 305-312

Slaughter, L. A., Soergel, D. und Rindfleisch, T. C., 'Semantic representation of consumer questions and physician answers' (2006) 75(7) *Int J Med Inform* 513-529

Stvilia, Besiki et al, 'Assessing Information Quality of a Community-Based Encyclopedia' (Arbeit präsentiert bei den Proceedings of the 2005 International Conference on Information Quality (MIT IQ Conference), Cambridge, MA, USA, 2005)

Voss, J. und Danowski, P., 'Bibliothek, Information und Dokumentation in der Wikipedia' (2004) 55(8) *Information, Wissenschaft & Praxis* 457-462

Wilson, Petra, 'How to find the good and avoid the bad or ugly: a short guide to tools for rating quality of health information on the internet' (2002) 324 *Bmj* 598-602

Wong, Chris, *XML, RDF and the Semantic Web: application to knowledge management* University of Technology, Sydney, 2003)

Zeng, Q. T. et al, 'Assisting consumer health information retrieval with query recommendations' (2006) 13(1) *J Am Med Inform Assoc* 80-90

7.2 URI/URL

Berners-Lee, Timothy et al, Frequently Asked Questions about RDF (2004)
<<http://www.w3.org/RDF/FAQ>> am 20.04.2007

Brickley, Dan, The truth about the semantic Web... (2007)
<<http://www.flickr.com/photos/danbri/428172848>> am 22.04.2007

Fox, Susanna Online Health Search 2006 (2006)
<http://www.pewinternet.org/pdfs/PIP_Online_Health_2006.pdf> am 03.03.2007

Fox, Susanna und Rainie, Lee, Vital decisions (2002)
<http://207.21.232.103/pdfs/PIP_Vital_Decisions_May2002.pdf> am 16.04.2007

Fox, Susanna und Rainie, Lee, The Online Healthcare Revolution: How the Web helps Americans take better care of themselves (1999)

<http://www.pewinternet.org/pdfs/PIP_Health_Report.pdf> am 03.03.2007

Kleinz, Torsten, Kritik und Unterstützung für Wikipedia (2007)

<<http://www.heise.de/newsticker/meldung/88359>> am 23.04.2007

O'Reilly, Tim, *Harnessing Collective Intelligence* (2006) O'Reilly Radar

<http://radar.oreilly.com/archives/2006/11/harnessing_coll.html> am 11.03.2007

O'Reilly, Tim, Web 2.0 Compact Definition: Trying Again (2006) O'Reilly Radar

<http://radar.oreilly.com/archives/2006/12/web_20_compact.html> am 02.03.2007

o.V., Alexa Top 500 Web Sites (2007) Alexa Internet, Inc.

<http://www.alexa.com/data/details/traffic_details?url=wikipedia.org> am 10.05.2007

o.V., Dublin Core (2007) <http://de.wikipedia.org/wiki/Dublin_Core> am 25.04.2007

o.V., Eurostat Leitseite Europäische Kommission

<http://epp.eurostat.ec.europa.eu/portal/page?_pageid=1090,30070682,1090_33076576&_dad=portal&_schema=PORTAL> am 11.03.2007

o.V., Internet Monitor Austria - Kommunitaktion und IT in Österreich (2006)

<http://www.integral.co.at/dlimages/AIM-C_4Quartal_2006.pdf> am 03.03.2007

o.V., NLM Fact Sheet SPECIALIST Lexicon

<<http://www.nlm.nih.gov/pubs/factsheets/umlslex.html>> am 25.04.2007

o.V., NLM Fact Sheet UMLS Metathesaurus

<<http://www.nlm.nih.gov/pubs/factsheets/umlsmeta.html>> am 25.04.2007

o.V., Resource Description Framework (2007)

<http://en.wikipedia.org/wiki/Resource_Description_Framework> am 22.04.2007

o.V., Seitenstatistik (2007) <<http://de.wikipedia.org/wiki/Spezial:Statistik>> am
03.05.2007

o.V., World Internet Users and Population Stats (2007) Miniwatts Marketing Group
<<http://www.internetworldstats.com/stats.htm>> am 18.04.2007