



FernUniversität in Hagen

Lehrgebiet Praktische Informatik VIII
Wissensbasierte Systeme

Bachelorarbeit

Vergleich von Ansätzen zur Nachrichtenvermittlung im Internet und Implementierung eines Systems in RDF

Verfasst von
Kathrin Dentler

Erster Prüfer
Prof. Dr. Christoph Beierle

Zweiter Prüfer
Prof. Dr. Stephan Lukosch

Abstract

In dieser Arbeit werden zunächst verschiedene Ansätze behandelt, die der Nachrichtenvermittlung im Internet dienen. Diese werden beschrieben und anhand einiger Kriterien miteinander verglichen.

Dann geht es um das Semantic Web, dessen Technologien das World Wide Web mit neuem Potential bereichern, indem sie die Möglichkeit bieten, Daten zur Verfügung zu stellen, deren SEMANTIK für unabhängige Programme verarbeitbar ist. So können Rechner in Zukunft ihre Kapazitäten besser im Internet einbringen und den Menschen Arbeit abnehmen, so dass diese sich auf das konzentrieren dürfen, was Rechner (noch) nicht leisten: mit Kreativität, Intuition und Inspiration fruchtbar zusammen zu arbeiten und Neues zu schaffen.

Danach wird ein modellhafter Ansatz für ein Globales Gehirn beschrieben. Ziel ist ein Ansatz der Nachrichtenvermittlung im Internet, die analog zu den Funktionsweisen des Gehirns aufgebaut ist. Dazu werden Funktionsweisen eines einzelnen Gehirns auf ein Globales Gehirn übertragen, in dem der einzelne Mensch wie eine Zelle im Gehirn in ein intelligentes Gesamtsystem eingebunden ist.

Das ist auch die Grundidee von OLI-it, einer Möglichkeit zur Nachrichtenvermittlung. Nachrichten werden von ihrem Urheber in einem hierarchisch strukturierten Wortraum markiert. Der Urheber beschreibt sich, den Inhalt der Nachricht und die Eigenschaften der gewünschten Empfänger. Auf der Suche nach Nachrichten werden Filterprofile erstellt, die mit komplementären Markierungen versehen sind. Durch den Abgleich beider Markierungen lässt sich eine hohe Trefferqualität erreichen. In dem nächsten Kapitel wird OLI-it anhand der entwickelten Kriterien beurteilt.

Zuletzt werden das Semantic Web und OLI-it zusammengeführt, indem die Technologien des Semantic Web für OLI-it genutzt werden. Es wird ein Client als Referenzimplementierung vorgestellt, der nur über Web Services mit OLI-it kommuniziert und dazu dient, neue Nachrichten und Filterprofile zu erstellen und abzugleichen.

Inhaltsverzeichnis

1	BESCHREIBUNG UND VERGLEICH VON ANSÄTZEN ZUR NACHRICHTENVERMITTLUNG IM INTERNET	1
1.1	Beschreibung der einzelnen Ansätze	1
1.1.1	E-Mail	1
1.1.2	Mailinglists	2
1.1.3	Usenet	2
1.1.4	CSCW	4
1.1.5	Chat	6
1.1.6	Marktplätze	7
1.1.7	Webforen	7
1.1.8	Wikis	8
1.1.9	Weblogs	9
1.2	Vergleich der Ansätze	10
1.2.1	Gemeinsamkeiten	10
1.2.2	Unterschiede	11
2	SEMANTIC WEB	18
2.1	Vision	18
2.2	Extensible Markup Language	20
2.2.1	Sprache	21
2.2.2	Struktur	21
2.2.3	Namespaces	22
2.3	Resource Description Framework	22
2.3.1	Datenmodell	23
2.3.2	RDF/XML	24
2.3.3	RDF Schema	25
2.3.4	Semantik	29
2.3.5	SPARQL Query Language	31
2.4	Web Ontology Language	32
2.4.1	Taxonomien	33
2.4.2	Tagging	33
2.4.3	Ontologien	34
2.4.4	OWL Vokabular	34
2.4.5	Untersprachen von OWL	37
2.5	Logik, Regeln und Beweise	38
2.6	Vertrauen	40

3	GLOBAL BRAIN	42
3.1	Das Gehirn	42
3.2	Die Vernetzung mehrerer Gehirne	45
4	BESCHREIBUNG VON OLI-IT	47
4.1	Kreislauf	47
4.1.1	Elemente	47
4.2	Wortraum	48
4.3	Vermittlungslogik	49
4.3.1	Markierung	49
4.3.2	Werte	50
4.3.3	Abgleich	52
4.4	Realisierung	52
5	BEURTEILUNG VON OLI-IT	53
5.1	Beurteilung von OLI-it anhand der vorgestellten Kriterien	53
5.2	Nachteile	53
5.3	Vorteile	54
6	RDF IMPLEMENTIERUNG VON OLI-IT	57
6.1	RDF Schema	57
6.2	RDF	59
6.3	Web Services	62
6.4	Verteilbarkeit	63
6.5	OLIClient	63
6.6	UniSonus	63
	Literaturverzeichnis	i
	Glossar	ii
	Stichwortverzeichnis	iv
	Abbildungsverzeichnis	v
	Erklärung	vi

1 Beschreibung und Vergleich von Ansätzen zur Nachrichtenvermittlung im Internet

Kommunikation basiert auf dem wechselseitigen Austausch von Nachrichten. Im Folgenden wird der Begriff Nachricht als eine Mitteilung von einem oder mehreren Urhebern an einen oder mehrere Empfänger verwendet. Mit einer Nachricht lassen sich beliebige Inhalte, wie beispielsweise Informationen, Anfragen, Fragen, Antworten oder Aussagen, ausdrücken. Damit werden auch Artikel in Wikis und Einträge in Weblogs als Nachrichten aufgefasst. Im Internet haben sich viele verschiedene Ansätze für die Vermittlung von Nachrichten und dementsprechend vielfältige Möglichkeiten der Kommunikation entwickelt. In diesem Kapitel werden zunächst die einzelnen Ansätze zur Nachrichtenvermittlung im Internet beschrieben und dann anhand einiger Kriterien miteinander verglichen.

1.1 Beschreibung der einzelnen Ansätze

In diesem Abschnitt werden die verschiedenen Ansätze der Nachrichtenvermittlung im Internet in der Reihenfolge ihrer Entstehung beschrieben. Es fällt auf, dass viele Ansätze schon lange vor der Erfindung des World Wide Web entwickelt wurden.

1.1.1 E-Mail

Nachdem das ARPANET¹ 1969 erstmals realisiert wurde, war E-Mail 1971 die erste und ist bis heute die meistgenutzte Möglichkeit der computervermittelten Kommunikation. E-Mail wurde nicht gezielt von den Entwicklern des ARPANET geplant und eingeführt, sondern eroberte das Netz aufgrund des Benutzerverhaltens.

Der Begriff E-Mail bezeichnet sowohl einen Dienst des Internet als auch eine einzelne E-Mail. Er stammt von der englischen Bezeichnung Electronic Mail und steht damit für elektronische Post. Im Vergleich zu Briefen werden E-Mails schneller und günstiger übermittelt. Da E-Mails im Klartext übertragen werden, gleichen sie jedoch eher einer Postkarte als einem verschlossenen Brief. Durch Verschlüsselungsverfahren wie PGP kann vertrauliche und authentische Kommunikation erreicht werden.

E-Mail bietet eine Möglichkeit für die Nachrichtenübermittlung zwischen einem Absender und gegebenenfalls mehreren Empfängern über ein Netzwerk. Mittels CC (Carbon Copy) und BCC (Blind Carbon Copy) können Kopien einer E-Mail zur Kenntnisnahme an weitere Empfänger vermittelt werden, wobei diese bei der Blind Carbon Copy für die anderen Empfänger nicht sichtbar sind. Absender und Empfänger werden durch ihre E-Mail Adressen eindeutig identifiziert. Eine E-Mail Adresse, wie sie für den Transport mittels SMTP (Simple Mail Transfer Protocol) verwendet wird, besteht aus einem lokalen Teil und einem globalen

¹ Wikipedia ARPANET: <http://en.wikipedia.org/wiki/Arpanet>

Teil, der die Domain bezeichnet. Diese beiden Teile sind durch das At-Zeichen (@) miteinander verbunden. E-Mails werden über ein E-Mail Programm oder eine Webschnittstelle (Webmail) gesendet und empfangen.

E-Mail substituiert teilweise die Kommunikation mittels Telefon, Post und Telefax, wird jedoch wegen ihrer geringen Kosten, hohen Vermittlungsgeschwindigkeit und der Möglichkeit, digitale Dokumente als Anhang mitzuschicken, auch zusätzlich zu anderen Medien verwendet. Da dem Absender die E-Mail Adressen der gewünschten Empfänger bekannt sein müssen, ist E-Mail eine relativ persönliche Form der Kommunikation, die vor allem für den privaten und organisatorischen Informationsaustausch verwendet wird.

1.1.2 Mailinglists

Die erste große Mailinglist war 1975 die Liste SF-LOVERS im ARPANET, in der immer mehr Forscher über Science Fiction diskutierten. Die ARPA sperre die Liste für einige Monate, da diese wenig mit Forschung zu tun hatte. Mit dem Argument, dass SF-LOVERS ein wichtiges Projekt zur Erforschung der Verwaltung und des Betriebs großer Mailinglists war, wurde die Liste wieder freigeschaltet. Tatsächlich musste das System wiederholt umgebaut werden, um dem stark ansteigenden Nachrichtenaufkommen gerecht zu werden.²

Eine Mailinglist ist eine Liste von E-Mail Adressen, die unter einer einzigen E-Mail Adresse zusammengefasst wird. Eine an die Listenadresse verschickte E-Mail wird an alle Mitglieder der Liste weitergeleitet. Da diese dem Absender nicht bekannt sein müssen, ist die Adressierung unpersönlicher als bei normalen E-Mails. Teilnehmer, die sich durch einen eigenen Beitrag zu erkennen gegeben haben, können jedoch von den anderen Teilnehmern durch die Verwendung der Absenderadresse direkt erreicht werden. Eine Mailinglist kann per E-Mail abonniert oder abbestellt werden. Normalerweise sind die Mitglieder einer Liste an einem bestimmten Thema interessiert, über das in der Mailinglist diskutiert wird. Mailinglists werden auch gerne für Informationen, Ankündigungen und Bekanntmachungen zu dem betreffenden Thema verwendet. Ein Beispiel sind die Mailinglists des W3C³, die auch im WWW veröffentlicht werden.

1.1.3 Usenet

Das Usenet entstand 1979 als Telefonverbindung zweier Unixrechner an zwei Universitäten. Es besteht aus Newsgroups, die hierarchisch in thematische Kategorien geordnet sind. In einer Newsgroup werden Nachrichten, so genannte Postings, veröffentlicht. Jeder User kann neue Themen beginnen und auf alle Nachrichten antworten. Dadurch entstehen Diskussionsbäume, die Threads genannt werden und chronologisch geordnet sind. Der

2 Wikipedia E-Mail: <http://de.wikipedia.org/wiki/E-Mail>

3 W3C Public Mailing List Archives: <http://lists.w3.org>

Sender eines Postings gibt seine bzw. eine E-Mail Adresse an und ermöglicht es damit dem Teilnehmerkreis, ihn bei Themen, die nicht die gesamte Newsgroup betreffen, direkt per E-Mail zu kontaktieren.

Postings werden meist über Newsreader empfangen und gesendet; an öffentlichen Newsgroups kann man auch über eine Webschnittstelle wie Google Groups⁴ teilnehmen. Die Nachrichten werden auf Newsservern gespeichert, die sich untereinander abgleichen. Dadurch entsteht eine dezentrale Architektur ohne zentrale Regelungsinstanz. Das im Usenet verwendete Protokoll zur Übertragung von Nachrichten ist seit 1985 NNTP (Network News Transfer Protocol).

Newsgroups dienen der Veröffentlichung von Bekanntmachungen, der Verbreitung von Informationen oder Angeboten, dem kooperativen Lösen von Problemen, dem Stellen und Beantworten konkreter Fragen sowie der Hilfe und sozialen Unterstützung in besonderen Lebenslagen.

In der langen Zeit, in der Newsgroups bestehen, hat sich eine eigenständige Kultur mit langer Tradition herausgebildet, die sich in der so genannten Netiquette widerspiegelt. Über die Netiquette wird in der Gruppe de.soc.netzkultur. umgangsformen diskutiert. Sie enthält allgemeine Empfehlungen für korrekte Umgangsformen im Usenet, die von den meisten Teilnehmern als sinnvoll erachtet werden. Autoren, die mehrfach gegen die Netiquette verstößen, werden in der Regel von anderen Teilnehmern verwarnt und können (falls dies keine Wirkung zeigt) mit Unterstützung der meisten Newsreader in ein so genanntes Killfile eingetragen werden, so dass zukünftige Postings dieser Absender nicht mehr angezeigt werden. Die Bekanntgabe dieses Vorgangs in der betreffenden Newsgroup wird im Netzjargon als plonken bezeichnet.

Ein weiteres Beispiel für die basisdemokratische Selbstregulierung und Organisation in Newsgroups sind die Regeln für die Einrichtung neuer Newsgroups in der de. Hierarchie. Diese sind nicht verpflichtend, werden aber allgemein befolgt.⁵ Eine Person, die eine neue Newsgroup einrichten möchte, muss zunächst ein RfD (Request for Discussion) per E-Mail an die Moderation der Newsgroup de.admin.news.announce schicken. Dieser Aufruf zur Diskussion wird nach Überprüfung durch die Moderation in die Newsgroup gepostet. Danach folgt eine mindestens zweiwöchige Diskussionsphase in der Newsgroup de.admin.news.groups, die solange andauert, bis Konsens erreicht wird. Dann wird ein CfV (Call for Vote) per E-Mail an die Moderation gesendet, der bei erfolgreicher Überprüfung wiederum in die Gruppe de.admin.news.announce gepostet wird. Mit der Veröffentlichung des Aufrufs zur Abstimmung beginnt die Abstimmung, deren Dauer im Vorhinein festgelegt

4 Google Groups: <http://groups.google.de>

5 Einrichtung neuer Newsgroups: <http://www.kirchwitz.de/~amk/dai/einrichtung>

sein muss; üblich sind drei Wochen bis ein Monat. Stimmen müssen bei dem aus dem CfV ersichtlichen Wahlleiter per E-Mail abgegeben werden. Eine Stimme ist nur dann gültig, wenn der Realname der wählenden Person angegeben wird, der nach der Wahl veröffentlicht wird. Die Wahlen sind also nicht geheim. Gewonnen ist eine Wahl, wenn nach der Abstimmung eine Zweidrittelmehrheit für die neue Newsgroup gestimmt hat, mindestens sechzig Ja-Stimmen abgegeben wurden und bis eine Woche nach dem Ende der Abstimmung keine Einsprüche eingegangen sind. Dann wird die neue Newsgroup eingerichtet.

1.1.4 CSCW

Der Begriff CSCW (Computer-Supported Cooperative Work) fiel erstmals im Jahr 1984, um das Thema eines Workshops zu beschreiben. Im Jahr 1986 fand die erste CSCW-Konferenz in den USA statt. Unter CSVW, also rechnerunterstützten kooperativen Arbeiten, versteht man den Einsatz von Software zur Unterstützung von Zusammenarbeit in Arbeitsgruppen oder Teams.⁶

CSCW dient damit nicht wie die anderen hier vorgestellten Ansätze der Vermittlung allgemeiner Nachrichten oder Informationen, sondern speziell der Unterstützung der Kooperation beziehungsweise Kollaboration, die immer der Kommunikation und Koordination bedarf. Werkzeuge zur Zusammenarbeit in Gruppen kann man daher in solche untergliedern, die die Kommunikation, die Koordination oder die Kooperation unterstützen. Diese lassen sich wiederum in synchrone und asynchrone Werkzeuge unterteilen [SSU01].

Systeme für die synchrone Kommunikation unterstützen die schriftliche Kommunikation durch die Vermittlung von Text und die mündliche Kommunikation durch die Vermittlung von Ton und bewegten Bildern. Audiovermittelte Kommunikation beinhaltet verbale und paraverbale Kommunikation wie Tonfall und Sprachtempo; videovermittelte Kommunikation beinhaltet darüber hinaus auch die nonverbale Kommunikation wie Mimik und Gestik. Videovermittelte Kommunikation gleicht daher der Face-to-Face Kommunikation am ehesten, wobei die Qualität eines Gespräches von Angesicht zu Angesicht bis heute unerreichbar ist und es in Zukunft schon wegen dem unterschiedlichen Ort der Kommunikationspartner, der diese Form der Kommunikation erst nötig macht, wohl auch bleibt. Schriftliche Kommunikation hat den Vorteil, dass kein technisches Zubehör wie Kopfhörer, Kameras oder Mikrofone benötigt wird. Werkzeuge zur Unterstützung der asynchronen Kommunikation sind beispielsweise E-Mail, Webforen und Newsgroups und werden in den weiteren Kapiteln behandelt.

6 GI Fachgruppe CSCW: <http://wwwfgcscw.in.tum.de/index.html>

Koordination ist das Abstimmen verschiedener Aktivitäten, die voneinander abhängig sein können, aufeinander. Komponenten, die die Koordination unterstützen, sind zum Beispiel Gruppenkalender, um Termine zu koordinieren, Workflowsysteme und Projektmanagementwerkzeuge, um strukturiert und unstrukturiert ablaufende Aktivitäten zentral zu koordinieren, To-Do Listen für die dezentrale Koordination sowie Tagesordnungen und Agenden, um Sitzungen zu koordinieren.

Synchrone kooperative Werkzeuge unterstützen die Kommunikation, Koordination und Kommunikation während der Zusammenarbeit. Die Kommunikation während der Zusammenarbeit wird neben den oben genannten Werkzeugen durch Anmerkungen und Zeiger erleichtert. Die Koordination der Zusammenarbeit wird vor allem durch Awareness ermöglicht. Awareness unterstützt die Gruppenwahrnehmung, vermittelt ein Gruppenbewusstsein und liefert dem einzelnen Benutzer idealerweise Informationen darüber, wer in der Vergangenheit was gemacht hat, wer momentan an welchen Dokumenten arbeitet und wer in der Zukunft welche Aufgaben hat. Die Kooperation umfasst die Bearbeitung gemeinsamer Dokumente wie beispielsweise Texte, Grafiken oder Hypertexte, gemeinsames Browsen in Webdokumenten und die gemeinsame Nutzung gemeinsamer Anwendungen. Zur Unterstützung der gemeinsamen Arbeit dient die Versionierung von Dokumenten, individuelle sowie gemeinsame Sichten auf die Dokumente, schnelle Rückmeldungen, Konsistenz und Fehlertoleranz. Kooperationsorientierte asynchrone Werkzeuge zeichnen sich dadurch aus, dass individuelle Arbeitsschritte der einzelnen Benutzer zu unterschiedlichen Zeiten erfolgen. Dokumente und Werkzeuge, um diese zu bearbeiten und zu strukturieren, liegen meist in gemeinsamen Arbeitsbereichen. Der Koordination der Zusammenarbeit dienen Metainformationen wie beispielsweise Anmerkungen.

Die einzelnen Werkzeuge werden je nach Bedarf zu domänenunabhängigen oder domänenspezifischen Technologiebündeln zusammengefasst. Generische Technologiebündel dienen beispielsweise dem Workflowmanagement, domänenspezifische sind auf ein bestimmtes Anwendungsbereich wie kooperatives Lehren und Lernen oder die kooperative Softwareerstellung hin optimiert.

Generische Mechanismen: Anpassbarkeit, kooperative Filtermechanismen,

äußere Umgebung hat Einfluss auf die Kommunikation -> gemeinsam genutzte Räume als Orte für die gemeinsame Bearbeitung von gemeinsam genutzten Material besser geeignet als verschachtelte Ordner -> intuitive Metapher

gestuftes Berechtigungskonzept, Zugriffsrechte für die Koordination, Zugriffskoordination

Für die Diskussion und Zusammenarbeit in geschlossenen Gruppen an der FernUniversität in Hagen eignet sich CURE⁷ (Collaborative Universal Remote Education), ein CSCL-Portal (Computer Supported Cooperative Learning). CURE bietet mehrere synchrone Tools wie DyCE zum Erstellen und Bearbeiten gemeinsamer Dokumente, einen gemeinsamen Text-Editor, ein gemeinsames Mind-Mapping Tool und ein gemeinsames Whiteboard. Die Gruppenbildung in CURE erfolgt mittels virtueller Räume und Schlüsseln mit Rechten für geschlossene Lernbereiche. Die einzelnen Räume, die aus mit einer wikiähnlichen Beschreibungssprache editierbaren Seiten bestehen, bieten weitere Funktionalitäten für die Zusammenarbeit wie Awareness (welche Nutzer befinden sich momentan im selben Raum), einen dauerhaft gespeicherten Chat und eine Threaded-Mail-Diskussion.

1.1.5 Chat

Chat (engl. to chat: plaudern, schwatzen) ermöglicht schriftliche Echtzeit-Kommunikation. Dabei werden die Nachrichten zeilenweise an die Kommunikationspartner vermittelt. Im Gegensatz zum Chat werden in MUDs (Multi User Dungeons) die Kommunikate zeichenweise übermittelt, was auch als Talk bezeichnet wird [Bec06]. Ein Problem dieser Form der Nachrichtenvermittlung sind Überlappungen bis hin zu zusammenhangslosen Textströmen, die durch unterschiedliche Tipp- und Übertragungsgeschwindigkeiten, hohe Teilnehmerzahlen und sich überschneidende Dialoge entstehen können.

Die älteste Form des Chat ist der Internet Relay Chat (IRC), der 1988 entwickelt und auf das Internet übertragen wurde. (Das ARPANET wurde 1982 nach der Adoption von TCP/IP (Transmission Control Protocol / Internet Protocol) zunehmend mit dem Namen Internet bezeichnet.) Der IRC setzt sich aus mehreren parallelen Netzwerken zusammen, die wiederum aus dezentralen, weltweit verteilten Servern bestehen. Auf die Netzwerke kann mit Clientsoftware in Form eigenständiger Clientprogramme oder Webclients zugegriffen werden. Im IRC wird in so genannten Channels, die jeder Teilnehmer nach Belieben neu erstellen kann, öffentlich gechattet. Channels sind Gesprächskanäle, die durch ihren Namen das Gesprächsthema vorgeben. IRC Programme zeigen dem Benutzer alle anderen Benutzer an, die sich im selben Channel befinden. Teilnehmer können sich für andere Benutzer unsichtbare Nachrichten zukommen lassen oder direkte Client-to-Client Verbindungen für private Gespräche und die Übertragung von Dokumenten aufzubauen. Der IRC bietet wie das Usenet einige Mechanismen für die Selbstregulierung. Normale Nutzer können sich gegenseitig ignorieren, Operatoren können gegen die Chatiquette verstößende Nutzer aus dem Chat ausschließen, sich gegenseitig zum Operator ernennen oder sich diesen Status nehmen. Der erste Operator ist derjenige, der den Channel eröffnet hat.

7 CURE: <http://teamwork.fernuni-hagen.de/CURE>

Eine weitere Form ist der Webchat, der dem IRC nachempfunden ist und deshalb oft ähnliche Funktionalitäten bietet. Allerdings ist ein Webchat auf die zugehörige Webpräsenz beschränkt und hat deshalb einen kleineren Teilnehmerkreis. Statt in Channels findet die Kommunikation in Chatrooms statt. Neben rein textbasierten Systemen existieren auch Chatsysteme mit zwei- oder dreidimensionalen grafischen Benutzeroberflächen und so genannten Avataren als Repräsentanten für Personen, die sich in virtuellen Räumen bewegen können.

Daneben gibt es das Instant Messaging, für das Messenger installiert werden müssen. Diese Programme bieten dem Benutzer oft Listen zur Anzeige bekannter Chatpartner, die gerade online und gesprächsbereit sind. Die Übertragung von Dokumenten ist so wie im IRC ein übliches Feature. Viele Messenger bieten darüber hinaus einen Audio Chat, der sich vor allem für private Diskussionen mit bekannten Personen in kleineren Gruppen eignet. Audio Chat über Messenger ist ein Vorläufer des Telefonierens über das Internet, auch Voice over IP genannt, das momentan dabei ist, den Markt zu erobern.

Chat eignet sich ebenso für die Kommunikation mit unbekannten als auch für die mit bekannten Menschen. Chatten kann sowohl eine Freizeitbeschäftigung als auch eine Form der synchronen Kommunikation während der Zusammenarbeit sein. Oft werden von Zeitschriften, dem Fernsehen oder Politikern Chat Events ins Leben gerufen, wobei feste Zeiten vorgegeben werden, zu denen mit bestimmten Persönlichkeiten gechattet werden kann. Hierbei handelt es sich meist um moderierte Webchats, das heißt, dass nicht alle Fragen oder Kommentare veröffentlicht werden.

1.1.6 Marktplätze

Das World Wide Web entstand im Jahr 1989 am CERN. Ab 1993 erhielt es durch den Browser Mosaic, der neben Text auch automatisch Graphiken anzeigen konnte, rasanten Auftrieb. Eine der ersten Anwendungen im Web waren ab 1995 Marktplätze, auf denen Nutzer Angebote und Gesuche für Käufe, Verkäufe oder die Partnervermittlung abgeben können. Dem Benutzer stehen meist von Experten konzipierte Eingabemasken für fest vorgegebene Kriterien zur Verfügung, nach denen gesucht und zwischen den Interessierten vermittelt werden kann.

1.1.7 Webforen

Webforen werden seit 1995 immer populärer. Ein Webforum ist eine Diskussionsplattform einer Webpräsenz und damit die erste Möglichkeit für Nutzer ohne eigene Homepage, themenbezogene Fragen oder Informationen ins Web zu stellen. Wie in einer Newsgroup entstehen Themen und Threads, die in einer Baumansicht dargestellt werden. Eng damit verwandt sind Boards, in denen der erste Beitrag das Thema vorgibt und der Beitrag selbst sowie die Antworten und Reaktionen auf diesen sequentiell als Liste dargestellt werden. Oft

findet man Mischformen aus Foren und Boards.⁸ Foren sind nicht wie Newsgroups global verteilt, sondern beziehen sich nur auf die entsprechende Webpräsenz und haben einen dementsprechend kleineren Benutzerkreis. Oft ist eine Anmeldung erforderlich, wenn man nicht nur lesen sondern auch schreiben möchte.

Ein aktuelles Beispiel ist Lycos iQ⁹, das als "menschliche Suchmaschine" vermarktet wird und ein Fragen- und Antwortensystem mit einer Experten Community verbindet. Dadurch entsteht ein soziales Netzwerk mit persönlichem Charakter.¹⁰ Lycos iQ nutzt Tagging, um Fragen zwischen den Experten zu vermitteln. So werden die Fragen mit Tags versehen und den Experten die Tags ihrer gestellten Fragen und der Fragen, auf die sie geantwortet haben, zugeordnet. Anhand dieser Tags werden Experten Fragen mit passenden Tags vorschlagen und Experten mit ähnlichen Interessensgebieten einander vorgestellt. Interessant ist hierbei, dass dem Benutzer ein Teil des Selektionsaufwands abgenommen werden kann, der durch die Masse der Nachrichten zwangsläufig entsteht.

1.1.8 Wikis

Wikis entstanden 1995 als Wissensmanagement-Tool. Der Name Wiki stammt von wikiwiki, dem hawaiianischen Wort für „schnell“.¹¹ Wikis sind kollaborative und dynamische Webseiten, die von jedem ohne technisches Hintergrundwissen online editiert werden können. Auf den zu den einzelnen Seiten gehörenden Diskussionsseiten wird idealerweise Konsens gefunden. Wikis realisieren damit eine ursprüngliche Idee von Tim Berners-Lee für das World Wide Web, Seiten nicht nur lesen, sondern auch direkt editieren zu können [BF00]. Der stärkste Einwand gegen das Konzept, dass jeder alles editieren kann, ist die Möglichkeit des Missbrauchs. Die Praxis zeigt jedoch, dass Missbräuche schnell entdeckt und behoben werden. Bestand hat nur, was von der Gemeinschaft akzeptiert wird. Als zusätzliche Sicherheit werden alle Änderungen und regelmäßig Versionen gespeichert, so dass kein großer Schaden entstehen kann. Neben Wikis, an denen jeder alles editieren darf, gibt es auch geschlossene Wikis.

In Wikis können durch typische vereinfachte Beschreibungssprachen Links auf andere Seiten erstellt werden, indem der Name der verlinkten Seite angegeben wird. Existiert die entsprechende Seite noch nicht, wird sie neu erstellt. Durch diese einfache Funktionsweise sind die einzelnen Seiten eines Wikis stärker untereinander verlinkt als normale Webseiten. Die Links sind zugleich die wichtigste Struktur der Wikis. Zudem gibt es meist auch Kategorien und Unterkategorien sowie Themenportale, um gewünschte Artikel zu finden.

8 Foren und Boards: <http://aktuell.de.selfhtml.org/artikel/gedanken/foren-boards/index.htm>

9 Lycos iQ: <http://iq.lycos.de>

10 Pressetext: <http://www.pressetext.de/pte.mc?pte=060222026>

11 Wikipedia Wiki: <http://de.wikipedia.org/wiki/Wiki>

Das momentan bedeutendste Wiki ist die Wikipedia¹², eine freie Enzyklopädie, deren Ziel es ist, die Gesamtheit des Wissens unserer Zeit zusammenzufassen und anzubieten. In der deutschen Version gibt es über 400.000 Artikel, die seit 2002 von einer unbekannten Anzahl freiwilliger Editoren und über 200.000 angemeldeten Benutzern verfasst wurden. Von diesen haben über 200 Administratorenrechte.

Eine Schwierigkeit des Prinzips, dass jeder alles ändern darf, ist die Konsensfindung. Dazu gibt es in der Wikipedia einige unumstößliche Richtlinien, die besagen, dass das Ziel die Schaffung einer Enzyklopädie ist, dass Inhalte neutral verfasst und verifizierbar sein sollen (Verzicht auch Primärrecherche), unter der GNU General Public License veröffentlicht werden und dass Mitglieder einen respektvollen Umgang ohne persönliche Angriffe miteinander pflegen sollen. Neben diesen Grundsätzen erfolgt die Organisation der Inhalte durch die gegenseitige Kontrolle und Korrektur der Mitglieder, die bei Konflikten in einem offenen Redaktionsprozess darüber entscheiden, welche Themen in welcher Form aufgenommen werden. Erscheint ein Thema beispielsweise als unangebracht, kann ein Löschantrag gestellt werden, über den offen diskutiert wird. Die Organisation der Wikipedia und ihrer Inhalte ist also immer ein sozialer Prozess der Selbstorganisation, der langwierige Diskussionen nicht ausschließt. Angemeldete Teilnehmer können sich durch gute Beiträge und überzeugende Argumente innerhalb der Community einen Ruf und Vertrauen erwerben und dadurch mehr Einfluss auf laufende Diskussionen gewinnen. Besonders engagierte Teilnehmer können zu Administratoren mit erweiterten Rechten gewählt werden, die zum Beispiel die Löschung von Seiten oder das Verhängen von IP-Sperren beinhalten. Diese haben allein wegen ihrer geringen Zahl besonderen Einfluss, den sie auch dafür verwenden sollten, ausufernde Diskussionen zu schlichten. Die oberste Instanz in der Wikipedia-Hierarchie ist die Betreiberorganisation Wikimedia Foundation.

1.1.9 Weblogs

Weblogs haben sich aus Webseiten entwickelt, die regelmäßig aktualisiert wurden. Ein Weblog, auch Blog genannt, ist eine Webseite, die wie ein Tagebuch periodisch neue Einträge enthält. Seit 1996 haben Nutzer (Blogger) die Möglichkeit, eigene Weblogs zu erstellen und mit Einträgen zu füllen. Veröffentlicht werden Berichte, Interessantes aus Streifzügen im Web, Bilder und Meinungen. Blogs dienen vor allem der Unterhaltung und der persönlichen Selbstdarstellung im Web, zunehmend aber auch der öffentlichen Meinungsbildung und dem Journalismus im Kleinen, wie es beispielsweise Blogs aus Krisengebieten zeigen. Weblogs unterstützen wie Wikis einfaches Publizieren ohne technische Vorkenntnisse. Jedoch arbeitet meist nur eine Person an den Inhalten. Diese können von den Lesern kommentiert werden, wobei die Kommentare wiederum

12 Wikipedia: <http://wikipedia.org>

veröffentlicht werden. Oft werden die Einträge automatisch als Newsfeeds zur Verfügung gestellt.

Bewertungen und Kommentare zu Informationen = Metainformation

1.2 Vergleich der Ansätze

Die Ansätze weisen viele Gemeinsamkeiten und einige Unterschiede auf, die im Folgenden beschrieben werden.

1.2.1 Gemeinsamkeiten

Die Kommunikation zwischen Menschen wurde durch das Internet um viele neue Möglichkeiten bereichert. Eine theoretisch unbegrenzte Anzahl von Nutzern kann unabhängig von Raum und Zeit in aller Welt bekannte oder fremde Menschen mit gemeinsamen Interessen, Eigenschaften oder sonstigen Gemeinsamkeiten erreichen. Dabei sind mehrere Kommunikationspartner möglich und üblich. Weiterhin kann jeder Empfänger ohne Medienbruch seine Antwort, sein Feedback oder ein Kommentar abgeben, wodurch Interaktion und Interaktivität ermöglicht werden. Oft bilden sich im Laufe der Zeit innerhalb der einzelnen Ansätze so genannte Communities mit eigenen Kulturen und eigenem Jargon heraus, in denen es beispielsweise üblich ist, dass die Mitglieder sich Duzen. Innerhalb einzelner Communities lässt sich eine große Hilfsbereitschaft erkennen, solange sich der Hilfesuchende an die Regeln der Gruppe (Netiquette) hält.

Internettechnologie

Die Basis für die Kommunikation im Internet ist die einheitliche TCP/IP Protokollfamilie, die den paketbasierten Datenaustausch zwischen und die Adressierung von Computern in Netzwerken standardisiert.

Schriftliche Kommunikation

Die Kommunikation verläuft bis auf den Audio Chat bei allen Anwendungen schriftlich ab. Das bedeutet, dass Faktoren wie Mimik, Gestik und Betonung, die bei der Kommunikation von Angesicht zu Angesicht eine große Rolle spielen, nicht vermittelt werden. Auch körperliche Anzeichen, die Rückschlüsse auf die soziale Identität, das Alter und Geschlecht der Kommunikationspartner zulassen, fehlen. Damit wird die Kommunikation anfälliger für Missverständnisse. Diesen kann durch die Verwendung so genannter Emoticons (Emotional Icons) und eine angemessene Ausdrucksweise entgegengewirkt werden.

Das Fehlen sozialer Hinweisreize hat auch positive Effekte (Kiesler & Sproull, 1992); z.B. werden Status- und Geschlechtsunterschiede reduziert (Egalisierung), was oft zu einer stärkeren Fokussierung auf die inhaltliche Aufgabe führt.

Digitale Form

Weiterhin ist allen diesen Formen gemeinsam, dass die Nachrichten in digitaler Form vorliegen, wodurch sie einfach, schnell und verlustfrei gespeichert, archiviert, weitergeleitet, verteilt, kopiert und editiert werden können.

1.2.2 Unterschiede

Die oben vorgestellten Ansätze der Nachrichtenvermittlung zu vergleichen, ist nicht einfach, da jeder Ansatz je nach Bedarf in den unterschiedlichsten Variationen eingesetzt wird. So können Webforen und Wikis in Intranets völlig von der Außenwelt abgeschirmt sein, viele Menschen gleichzeitig an einem Blog arbeiten oder nur einer an einem Wiki. Im Zweifelsfall werden die Anwendungen deshalb so verglichen, wie sie oben beschrieben sind und ursprünglich konzipiert wurden.

Die beiden häufigsten Kriterien, nach denen die verschiedenen Ansätze kategorisiert werden, sind die Einteilung in synchrone und asynchrone Kommunikation und die Anzahl der beteiligten Kommunikationspartner. Bei genauerer Betrachtung lassen sich jedoch viele weitere Unterschiede und damit Kriterien für die Abgrenzung der einzelnen Ansätze feststellen.

Synchrone oder asynchrone Kommunikation

Synchronität (syn: mit, gemeinsam; chronos: Zeit) bedeutet Gleichzeitigkeit. Von den vorgestellten Ansätzen der Nachrichtenvermittlung ermöglicht nur der Chat eine synchrone Kommunikation. Durch nicht garantierte schnelle Übertragungs-, Lauf-, Veröffentlichungs- und Zugriffszeiten kann die Kommunikation in den anderen Anwendungen quasi-synchron werden, wenn die Beteiligten jeweils unmittelbar antworten.

Sowohl die synchronen als auch die asynchronen Formen der Kommunikation haben jeweils Vor- und Nachteile. Synchrone Kommunikation ist direkt und erhöht für den Urheber einer Nachricht die Wahrscheinlichkeit einer zeitnahen Antwort bzw. Reaktion, erfordert aber die gleichzeitige Kommunikation der Beteiligten. Bei zu vielen Teilnehmern ergeben sich konfuse Gesprächsverläufe. Bei der asynchronen Kommunikation müssen die Empfänger einer Nachricht zu dem Zeitpunkt des Versands nicht online sein, sondern können die Nachricht zu einem für sie passenden Zeitpunkt lesen. Der Nachteil hierbei ist wiederum, dass der Urheber der Nachricht nicht von einer schnellen Antwort ausgehen kann.

Asynchron: man muss keinen Termin vereinbaren

Anzahl der Urheber und Empfänger einer Nachricht

Das zweite Kriterium ist die Anzahl der Urheber und der Empfänger einer Nachricht. In vielen Systematisierungen wie beispielsweise [MO96] werden die Kommunikation per E-Mail

als one-to-one und die in Mailinglists, Newsgroups und Webforen als many-to-many Kommunikationsmodi kategorisiert, da bei diesen Kommunikationsformen viele Urheber viele Nachrichten verfassen. Global betrachtet verfassen jedoch auch viele Urheber viele E-Mails, weshalb im Folgenden ein nachrichtenzentrierter Ansatz zur Kategorisierung der Ansätze gewählt wird.

Das Wiki ist die einzige Anwendung, in der mehrere Urheber gemeinsam an einem Inhalt arbeiten; bei allen anderen Ansätzen arbeitet in der Regel nur eine Person an dem Inhalt der Nachricht. Bei den meisten Kommunikationsformen geht die Nachricht an mehrere Personen. Ausnahmen sind E-Mails und Chat. E-Mails werden normalerweise nur an eine Person geschickt, können aber auch an mehrere Empfänger gerichtet sein. Gechattet wird normalerweise in einer Gruppe mit mehreren Personen, es ist aber auch möglich, einen abgegrenzten Chat zu zweit einzurichten.

Wenn viele Urheber an einem Inhalt arbeiten, hat das den Vorteil, dass schneller und mehr Inhalt generiert werden kann, als es einer einzelnen Person möglich wäre. Auf der anderen Seite muss die Qualität des Inhaltes sichergestellt und Konsens gefunden werden. Das dies möglich ist, zeigt das Beispiel der Wikipedia. Im Gegensatz dazu ist ein einzelner Urheber alleine und ohne Reibungsverluste für seine Inhalte verantwortlich. Mehrere Empfänger weisen auf eine unpersönlichen Adressierung oder eine Gruppe hin, ein Empfänger auf persönliche Kommunikation.

Zusammensetzung des Empfängerkreises

In allen Kommunikationsformen kann eine Nachricht an mehrere Empfänger adressiert sein. Dabei lässt sich unterscheiden, wie sich der Empfängerkreis zusammensetzt, also ob er bekannt, offen oder geschlossen ist.

Eine E-Mail kann nur an Empfänger vermittelt werden, deren E-Mail Adressen dem Absender bekannt sind. Eine Nachricht an eine Mailinglist wird an alle eingetragenen Empfänger weitergeleitet; diese müssen dem Absender nicht zwangsläufig bekannt sein. Einige Mailinglists sind gegenüber der Öffentlichkeit geschlossen, in andere darf sich jeder als Empfänger eintragen. Nur eingetragene Mitglieder bekommen die E-Mail übermittelt, wodurch der Empfängerkreis relativ geschlossen ist. Daneben werden einige Mailinglists wie beispielsweise die des W3C im Web veröffentlicht und sind damit für jeden einsehbar und offen. Obwohl es geschlossene Newsgroups gibt, die eine Authentifizierung erfordern oder deren Postings nur auf bestimmten Newssevern gehalten werden, ist das Usenet prinzipiell so angelegt, dass jeder alle Postings lesen kann. Chat gibt es in geschlossener und öffentlicher Form, so dass entweder nur die beteiligten Chatter die Nachrichten ihrer Kommunikationspartner empfangen können oder jeder, der sich zu dem gleichen Zeitpunkt in demselben Channel befindet. Nachrichten aus Webforen, Wikis und Weblogs können in

der Regel von allen gelesen werden. Auch Marktplätze sind normalerweise so konzipiert, dass der Empfängerkreis möglichst groß ist und somit jeder alles lesen darf.

Die öffentlichen Formen der Kommunikation finden mit, aber vor allem vor Fremden statt. So gibt es immer Empfänger, die sich nicht aktiv an der Kommunikation beteiligen, sondern nur lesen. Diese werden im Netzjargon als Lurker (engl. to lurk: lauern, herumschleichen) bezeichnet. Der negative Beiklang dieses Begriffes stammt daher, dass es unangenehm sein kann, vor Menschen, die nichts über sich selbst preisgeben, Nachrichten zu veröffentlichen. Dies passiert vor allem im Chat, da dort die Gegenwart der passiven Teilnehmer sichtbar ist. Im Gegensatz dazu erfährt der Urheber von Beiträgen in Mailinglists, Newsgroups und Webforen nichts von passiven Lesern, kann aber von ihrer Existenz ausgehen. In anderen Anwendungen wie Marktplätzen, Wikis und Weblogs, die gerade der Veröffentlichung von Informationen dienen, sind nur lesende Teilnehmer der Normalfall und explizit erwünscht.

Identifizierbarkeit des Urhebers durch die Kommunikationspartner

Jeder Rechner im Internet besitzt eine IP-Adresse, unter der er eindeutig identifizierbar ist. Das ist notwendig, damit der Rechner Datenpakete versenden und vor allem empfangen kann. Ein Server, der HTTP-Anfragen von Clients, die unter anderem die IP-Adressen der Clients enthalten, empfängt und verarbeitet, kann diese grundsätzlich mitprotokollieren. Dadurch kann der Anbieter eines Angebots beispielsweise feststellen, welche Seiten am häufigsten aufgerufen werden. IP-Adressen lassen Rückschlüsse auf den Internet Service Provider zu, der dem Benutzer den Zugang zum Internet ermöglicht.

Momentan gilt in der deutschen Rechtssprechung der Erforderlichkeitsgrundsatz, der besagt, dass personenbezogene Daten grundsätzlich nur gespeichert werden dürfen, wenn dies zu einem bestimmten, gesetzlich zugelassenen Zweck (wie zum Beispiel für Abrechnungen) erforderlich ist. In der Praxis speichern viele Internet Service Provider die Verkehrsdaten und die Zuordnung zwischen Nutzern und IP-Adressen für einige Zeit, auch wenn diese Daten nicht für Abrechnungszwecke erforderlich sind, und sind im Ernstfall einer Straftat zu der Herausgabe der Identität des Benutzers an Ermittlungsbehörden verpflichtet.

Benutzer im Internet sind daher nie wirklich anonym. Um dennoch Anonymität herzustellen, können Anfragen beispielsweise über einen Proxy Server geleitet werden, so dass die IP Adresse des Absenders verschleiert wird. Hierbei besteht die Gefahr, dass der Betreiber dieses Servers die Informationen zusammensetzen und herausgeben kann.

Eine sicherere Variante sind Anonymizer wie beispielsweise JAP¹³. Mit dieser Technologie teilen sich viele Benutzer eine IP-Adresse, so dass Server die Identität der Clients nicht

13 JAP: <http://anon.inf.tu-dresden.de>

feststellen können. Der JAP-Client dient als Proxy auf dem lokalen Rechner und baut eine verschlüsselte Verbindung zu einem Mix-Server auf, der die Anfrage über verschlüsselte Kommunikationsverbindungen und mehrere Zwischenstationen an den Ziel-Server und das Ergebnis der Anfrage auf dem gleichen Weg zurück liefert. Diese Zwischenstationen werden von unabhängigen Institutionen betrieben, die sich verpflichtet haben, weder Log-Files zu speichern noch Daten mit anderen Betreibern austauschen. Nachteil dieser Variante ist, dass das Surfen im Internet deutlich langsamer wird.

Für die Anonymisierung von E-Mails und Newsgroup Postings eignen sich Remailer, die E-Mails über eine oder mehrere Stationen verschlüsselt weiterleiten und dabei Informationen, die Rückschlüsse auf den Absender zulassen, entfernen. Der Nachteil dieses Verfahrens sind die zeitliche Verzögerung sowie der mögliche Verlust dieser E-Mails.

In der Regel fühlen sich Benutzer im Internet relativ anonym und nur wenige verwenden die oben genannten Methoden zur Anonymisierung. Abgesehen von diesen Methoden kann jeder Benutzer selbst bestimmen, wie viel er von seiner Identität freigibt. Viele Benutzer geben Hinweise auf ihre tatsächliche Persönlichkeit, zum Beispiel über Signaturen oder Angabe ihrer persönlichen Homepage.

Bei E-Mail (und damit auch Mailinglists) und Newsgroups authentifiziert sich der Urheber mit seiner E-Mail Adresse, die, wenn sie von Institutionen und Unternehmen stammt, meist Rückschlüsse auf die Person, der die Adresse gehört, zulässt. Vorname.Nachname -> Adressen können von externen „erraten“ werden.

Da E-Mail viel für die Kommunikation mit bekannten Personen verwendet wird, ist darüber hinaus die Angabe des realen Namens üblich. Die Usenet Netiquette schlägt es den Benutzern ausdrücklich vor, ihre echten Namen anzugeben, soweit es sich nicht um sensible Themen handelt. Bei allen anderen Anwendungen gibt der Benutzer normalerweise nur gegenüber dem Anbieter seine E-Mail Adresse an. Empfänger der Nachrichten sehen nur den frei wählbaren Benutzernamen. Benutzernamen und Pseudonyme dienen der Selbstrepräsentation und Wiedererkennung, es ist nicht unüblich, dass Nutzer ihren echten Namen als Benutzernamen angeben.

Gechattet wird nicht unter dem echten Namen, sondern unter frei wählbaren Pseudonymen, die im Chat als Nicknames bezeichnet werden. Bei der Kommunikation mit fremden Menschen entsteht dadurch eine gewisse Unverbindlichkeit und Verringerung von Hemmungen, die für das Spiel mit der eigenen Identität, aber auch für unsoziales Verhalten (Flaming) genutzt werden kann. So ist es beispielsweise möglich, ein anderes Geschlecht vorzugeben (Gender Switching) oder die eigene Person zu beschönigen.

Virtuelle Identität

Webanwendung oder eigenständiger Dienst

Ein weiterer Unterschied ist, dass E-Mail (und damit auch Mailinglists), das Usenet, der IRC und Instant Messaging eigene Dienste neben dem WWW sind, während die anderen Anwendungen auf dem WWW beruhen, also Webanwendungen sind. Für E-Mail, das Usenet und den IRC Chat existieren Webschnittstellen.

Zentrale oder dezentrale Architektur

Manche Anwendungen, wie zum Beispiel alle Webanwendungen, laufen zentral über einen Server. E-Mails sind dezentral über mehrere Server verteilt, Beiträge in Mailinglists sammeln sich auf dem Server der Verteiladresse und werden dann weitergeleitet. Das Usenet hat eine komplett dezentrale Architektur. Webchats gehen zentral über einen Server, der IRC über ein dezentrales Netzwerk von Servern. Diese Strukturen wirken sich darauf aus, wie demokratisch die Kommunikation im Allgemeinen abläuft, wie anfällig sie für Zensur ist und wie anonym die Teilnehmer sind.

Speicherung der Nachrichten auf dem eigenen Rechner

Eine weitere Unterscheidungsmöglichkeit für die einzelnen Anwendungen ist, ob die Nachrichten auf dem eigenen Rechner gespeichert werden, oder ob man die Anwendung besuchen muss, um neueste Nachrichten zu lesen. Es ist zu erkennen, dass nur E-Mails, Mailinglist- und Newsgroup Postings auf dem eigenen Rechner gespeichert werden. Newsgroup Postings verbleiben im Gegensatz zu E-Mails nach dem Speichern auf dem Server. Durch die Verwendung von IMAP (Internet Message Access Protocol) oder Webmail kann man jedoch auch auf E-Mails zugreifen, die auf einem Mailserver liegen.

Das WWW ist grundsätzlich ein Pull-Medium, das bedeutet, dass die Daten von Webanwendungen auf dem zentralen Server dieser Anwendungen liegen und man diese besuchen muss, um neue Nachrichten zu lesen. Um dies zu vermeiden gibt es zwei Lösungsansätze, die für alle Webanwendungen eingesetzt werden können. Webforen und Marktplätze informieren den Benutzer gerne per E-Mail über neueste Einträge. Wikis und Blogs bieten Newsfeeds, um Interessierte über Änderungen oder neue Einträge auf dem Laufenden zu halten. Damit lassen sich alle Nachrichten auch offline lesen. Offline schreiben kann man (ohne Hilfsbrücken wie einen Texteditor) nur E-Mails, Mailinglist Beiträge und Newsgroup Postings.

Der Chat lässt sich nur per Protokoll auf dem eigenen Rechner speichern, das sich in der Regel nur in Echtzeit erstellen lässt. Bei der Masse der Nachrichten wäre es nicht sinnvoll, Zugang zu allen Nachrichten zu haben. Messenger bieten oft die Möglichkeit, abwesenden Teilnehmern persönliche Nachrichten zu hinterlassen.

Anwendungsmodell aus der realen Welt oder neu

Zudem gibt es Anwendungsmodelle aus der realen Welt, die für das Internet übernommen wurden, und neue Modelle, die erst durch das Internet möglich wurden. Eine E-Mail ist vergleichbar mit einem Brief. Marktplätze ersetzen Flohmärkte und spezialisierte Händler. Audio Chats gleichen Telefonaten. Auch früher haben mehrere Autoren an Enzyklopädien gearbeitet, mit dem Unterschied, dass diese nicht anonym waren, da sie Verantwortung für ihre Artikel übernehmen mussten. Blogger bezeichnen sich gerne als Journalisten, die ohne ihr Blog nicht die Möglichkeit hätten, Inhalte einfach, schnell und günstig zu publizieren. Modelle des Usenet, Chat und der Webforen sind in ihrer Form in der realen Welt nicht zu finden.

Inhalt der Nachrichten

Zudem unterscheiden sich die Anwendungen wenigstens grob in ihrem Inhalt. Mailinglists, das Usenet und Foren dienen der Diskussion. Inhalte von E-Mails und des Chats ergeben sich aus den individuellen Interessen der Benutzer. Wikis bieten gesammelte Information und Blogs subjektive Einträge. Auf Marktplätzen findet man Angebote und Gesuche.

Moderierte oder unmoderierte Kommunikation

Eine weitere Unterscheidungsmöglichkeit ist, ob die Kommunikation moderiert oder unmoderiert ist. Bis auf E-Mails lassen sich alle Ansätze moderieren. Moderatoren können dafür verantwortlich sein, lenkend in Diskussionen einzugreifen, auf Fragen zu antworten, oder Nachrichten zu überprüfen, freizuschalten und gegebenenfalls zu löschen. Dabei können Moderatoren von dem Verantwortlichen der Anwendung bestimmt oder von den Beteiligten frei gewählt werden.

Nachträgliche Editierbarkeit der Nachrichten

Bei einigen Ansätzen lassen sich Nachrichten vom Urheber nachträglich editieren, bei anderen Anwendungen, wie beispielsweise E-Mail, Mailinglists und im Chat, sind die Nachrichten endgültig. Newsgroup Postings kann man nachträglich per Cancel Message löschen, was allerdings nicht alle Server unterstützen. Zudem sollte ein wirkungsvolles Löschen eines Postings relativ schnell nach dem Absenden geschehen, bevor die Leser der Gruppe das Posting auf ihrer Festplatte gespeichert haben. Zudem ist es möglich, einen Fremdcancel zu verschicken, also eine Cancel Message, die nicht von dem Urheber des zu löschen Postings stammt. Dies ist zwar offiziell nicht erlaubt, wird aber in einigen Fällen toleriert, zum Beispiel wenn die Cancel Message von dem Geschädigten eines Postings mit falschem Absender stammt. Analog dazu können mit Supersedes Messages Postings überschrieben werden.

Nachfolgend eine kleine Tabelle zur Gesamtübersicht:

Ansatz / Kriterium	E-Mail	Mailinglist	Usenet	Chat	Marktplätze	Webforen	Wikis	Weblogs
synchron / asynchron	synchron	synchron	synchron	asynchron	synchron	synchron	synchron	synchron
Kommunikationspartner	one-to-one / one-to-many	one-to-many	one-to-many	one-to-one / one-to-many	one-to-many	one-to-many	many-to-many	one-to-many
Empfängerkreis	bekannt	geschlossen / offen	geschlossen / offen	geschlossen / offen	offen	offen	offen	offen
Anonymität	E-Mail Adresse	E-Mail Adresse	E-Mail Adresse / Name	Benutzername	Benutzername	Benutzername	Benutzername	Benutzername
Web / Dienst	Dienst	Dienst (E-Mail)	Dienst	Dienst	Web	Web	Web	Web
Anwendungsmodell aus der realen Welt / neu	Brief	Rundbrief	neu	neu	Marktplatz / Händler	neu	Redaktion	Journalist
Speicherung der Nachrichten	ja	ja	ja	nein	nein	nein	nein	nein
Inhalt der Nachrichten	offen	Diskussion	Diskussion	offen	Angebote / Gesuche	Diskussion	gesammelte Information	subjektive Einträge
moderierbare Kommunikation	nein	ja	ja	ja	ja	ja	ja	ja
nachträgliche Editierbarkeit	nein	nein	frühes Löschen möglich	nein	ja	ja	ja	ja

2 Semantic Web

The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation.¹⁴

Das Semantic Web ist eine von vielen Aktivitäten des World Wide Web Consortium (W3C), dessen Mission es ist, interoperable Technologien zu entwickeln, um das Web zu seinem vollen Potential zu führen. Vorsitzender des W3C ist Tim Berners-Lee, der Erfinder des World Wide Web. Das W3C ist ein internationales Konsortium, in dem Mitgliedsorganisationen gemeinsam De-facto-Standards entwickeln, die auch als Recommendations bekannt sind.

Laut W3C¹⁵ wird das Web sein volles Potential erst dann erreichen, wenn dezentral erstellte Daten sowohl von unabhängig voneinander entworfenen Anwendungsprogrammen als auch von Menschen gemeinsam genutzt und verarbeitet werden können. Dies kann nur über offene Standards ermöglicht werden. Das Semantic Web ist eine Vision: die Idee von Daten im Web, die so definiert und miteinander verknüpft sind, dass sie von Computern nicht nur zu Darstellungszwecken, sondern auch zur Automatisierung, Integration und Wiederverwendung genutzt werden können.

2.1 Vision

Heute werden Inhalte im Web von Menschen für Menschen dargestellt. Liegen in der Zukunft Daten in einer Form vor, in der sich ihre Bedeutung auch maschinell verarbeiten lässt, erleichtert das die Zusammenarbeit zwischen Mensch und Maschine.

Ein Beispiel hierfür ist ein Softwareagent, den man sich als Schnittstelle zwischen dem Benutzer und dem Semantic Web vorstellen kann. Er übernimmt wie ein persönlicher Assistent diverse Aufgaben. Ein Beispiel ist die Planung einer Dienstreise: Flug, Hotel und Mietwagen sollen gebucht, Termine während der Reise vereinbart werden. Der Agent stellt mehrere Alternativen zusammen und der Benutzer kann die beste auswählen oder weitere Bedingungen stellen. Dabei lernt der Agent immer mehr über die persönlichen Interessen und Vorlieben seines Benutzers.

Auch bei der Suche im Web kann ein Agent behilflich sein: heute liefern Suchmaschinen Links als Treffer, die man selbst durchsuchen muss, um die gewünschte Information zu finden. Ein Agent hingegen kann Antworten auf Suchanfragen aus mehreren Dokumenten ermitteln. Bei zu vielen oder zu wenigen Treffern kann die Suche automatisch angepasst werden.

14 Tim Berners-Lee, James Hendler, Ora Lassila, The Semantic Web, Scientific American, Mai 2001: <http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21>

15 W3C Semantic Web Activity: <http://www.w3.org/2001/sw>

Um derartige Aufgaben zu erfüllen, müssen sich Agenten nicht nur mit dem Benutzer, sondern auch untereinander austauschen. Dazu werden Nachrichten mit Web Services übermittelt. Herausforderungen hierbei sind das Finden, Vermitteln und Zusammensetzen der benötigten Web Services. Ein anderer wichtiger Punkt ist, dass Agenten von bestimmten Konzepten der Anwendungsdomäne zumindest teilweise das gleiche Modell besitzen sollten. Hier haben Ontologien und deren Mapping, also ihre Abbildung aufeinander, eine große Bedeutung.

Der Ansatz des Semantic Web ist es, Sprachen zu entwickeln, mit denen Informationen in eine Form gebracht werden, die auch von Maschinen verarbeitet werden kann, und nicht, wie in der Künstlichen Intelligenz, Maschinen so zu programmieren, dass sie sich wie Menschen verhalten.¹⁶

Um die Vision des Semantic Web zu realisieren, werden Daten und Metadaten im Web mit Sprachen wie XML, RDF/S (RDF und RDF Schema) und OWL, einer Ontologiesprache, beschrieben. Metadaten sind Daten über Daten, die dazu genutzt werden, Ressourcen zu identifizieren, zu beschreiben oder zu lokalisieren.

Daten werden durch Semantik zu Informationen und Informationen werden durch Verknüpfungen zu Wissen, so dass das Internet durch das Semantic Web zu einer weltweiten Wissensbasis erweitert werden kann. Das Potential des Semantic Web liegt vielleicht gerade in dem, was wir uns heute noch nicht vorstellen können. Schon absehbar ist jedoch eine Evolution des menschlichen Wissens, wenn das Internet statt "Information at Your Fingertips" in Zukunft "Knowledge at Your Fingertips" bietet.

16 Semantic Web Road Map: <http://www.w3.org/DesignIssues/Semantic.html>

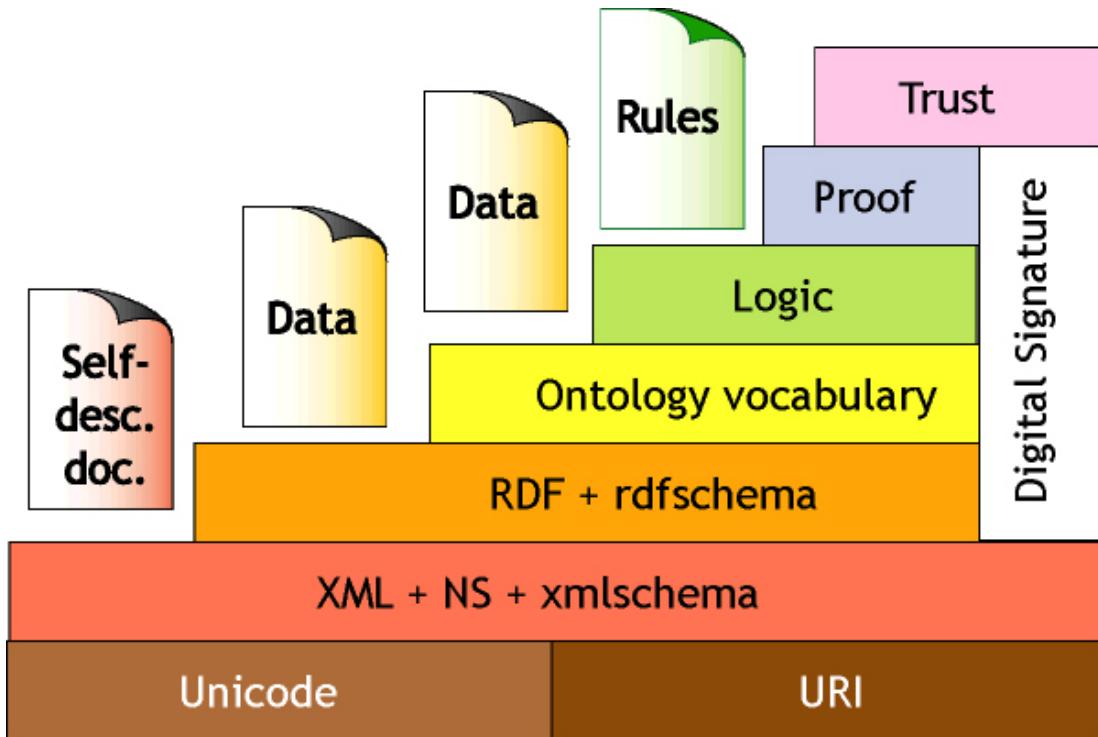


Abbildung 1: Die Schichten des Semantic Web¹⁷

Abbildung 1 zeigt die Schichten des Semantic Web, wobei höhere Schichten auf tiefere aufbauen und sie mitnutzen. Im Folgenden wird auf die einzelnen Schichten genauer eingegangen. Die oberen drei Schichten befinden sich noch in der Entwicklung und werden deshalb nur kurz beschrieben.

2.2 Extensible Markup Language

XML¹⁸, die Extensible Markup Language, ist eine einfache, sehr flexible Markup Sprache, die ihre Wurzeln in SGML, der Standard Generalized Markup Language, hat. XML bringt Struktur in Dokumente, womit syntaktische Interoperabilität unterstützt wird, die die Zusammenarbeit mehrerer Systeme ermöglicht. Deshalb spielt XML eine immer wichtigere Rolle für den Austausch von Daten im Web, zum Beispiel über Web Services, die in B2B Anwendungen schon eingesetzt werden. Dies ist ein großer Schritt in Richtung des Semantic Web, da XML die Grundlage für weitere Schichten wie RDF/S und OWL bildet.

Ein Web Service¹⁹ ist ein Softwaresystem, das Interaktion mit Programmen unterstützt. Die angebotenen Methoden eines Web Services und ihre Parameter werden in der Web Service Definition Language (WSDL) beschrieben. Andere Systeme interagieren mit dem Web Service gemäß dieser Beschreibung über Simple Object Access Protocol (SOAP)

17 Architekturdiagramm von Tim Berners-Lee: <http://www.w3.org/2000/Talks/1206-xml2k-tbl/slides10-0.html>

18 Extensible Markup Language (XML): <http://www.w3.org/XML>

19 W3C Web Services Activity: <http://www.w3.org/2002/ws>

Nachrichten, die typischerweise mittels HTTP übertragen werden. Ein Verzeichnisdienst zur Registrierung von Web Services ist Universal Description, Discovery and Integration (UDDI). Durch diesen Verzeichnisdienst wird das dynamische Finden der Services anhand von Taxonomien ermöglicht. Durch die Benutzung weit verbreiteter Internetstandards sind Web Services unabhängig von Plattformen, Programmiersprachen und Protokollen und ermöglichen eine offene und flexible Architektur für verteilte Anwendungen.

2.2.1 Sprache

XML ist eine Metasprache, die es erlaubt, eigene Formate zu entwickeln und Dokumente in diesen Formaten zu erstellen. Solche Dokumente bestehen aus Daten und Markup in der Form von Tags. Das Markup ermöglicht es Programmen, Elemente bzw. Inhalte richtig zu interpretieren.

Ein Element besteht aus einem öffnenden Tag, dem Inhalt und einem schließenden Tag, z.B.: <University>FernUniversität</University>

In einem öffnenden Tag können Attribute als Name-Wert-Paare angegeben werden:
<University Location="Hagen">FernUniversität</University>

XML enthält keine Angaben bezüglich der Darstellungsform der Daten. Durch die Verwendung verschiedener Stylesheets können XML Dokumente für unterschiedliche Anzeigemedien und Geräte beliebig transformiert und dargestellt werden.

2.2.2 Struktur

Ein XML Dokument heißt wohlgeformt, wenn alle syntaktischen Bedingungen erfüllt sind. Zum Beispiel muss zu jedem öffnenden Tag ein passendes schließendes Tag vorhanden sein und die Elemente müssen richtig, also ohne Überschneidungen, ineinander geschachtelt sein.

Die Struktur eines XML Dokumentes kann durch einen Baum dargestellt werden.²⁰ Jedem Element des Dokumentes und jedem Elementtext wird ein Knoten zugeordnet. Das Dokumentelement ist die Wurzel des XML Baumes. Ist ein Knoten mit einem Elementinhalt beschriftet, wird er als Textknoten bezeichnet. Alle anderen Knoten sind Elementknoten.

Für die Kommunikation zwischen verschiedenen Programmen wird ein gemeinsames Vokabular definiert, also die Element- und Attributnamen, die verwendet werden können. Außerdem wird die Dokumentstruktur festgelegt, indem bestimmt wird, welche Elemente und Attribute innerhalb anderer vorkommen können.

20 XML: http://lectures.informatik.uni-freiburg.de/lectures/DBTheorie/2005SS/Slides/01_03_dbT_XML.pdf

Um die Struktur von XML Dokumenten festzulegen, gibt es zwei Möglichkeiten: die ältere DTD (Document Type Definition), die als erweiterte Backus-Naur-Form interpretiert werden kann, und XML Schema.

Der Vorteil von XML Schema im Vergleich zu DTD ist zum einen, dass die SYNTAX auf XML selbst basiert, und zum anderen, dass XML Schemata wiederverwendet und neu definiert werden können. XML Schema stellt eine Menge von integrierten Standarddatentypen zur Verfügung und bietet die Möglichkeit, neue Datentypen zu erstellen, indem existierende erweitert oder mit strenger Bedingungen versehen werden.

Ein XML Dokument ist gültig, wenn es wohlgeformt ist und Information bezüglich seiner Struktur besitzt und erfüllt.

Die Semantik von XML Dokumenten und Schemata kann von Programmen nicht automatisch erschlossen werden, sondern muss für jede Anwendungsdomäne sinngemäß implementiert werden.

2.2.3 Namespaces

In einem XML Dokument lassen sich mehr als eine DTD oder ein XML Schema verwenden, wodurch jedoch Namenskonflikte auftreten können. Um dies zu vermeiden, werden die entsprechenden Quellen durch Präfixe unterschieden. Eine Namespace Deklaration hat die Form `xmlns:prefix="location"`. Namen aus diesem Namespace können im Dokument mit `prefix:name` bezeichnet werden, also mit einem XML qualified name, verkürzt QName. Ein Beispiel hierfür ist das Präfix `rdf`, das für das RDF Vokabular mit dem Namespace URI <http://www.w3.org/1999/02/22-rdf-syntax-ns#> steht.

URIs (Uniform Resource Identifiers) sind eindeutige Bezeichner für Ressourcen im Web. Im Gegensatz zu URLs (Uniform Resource Locators) beschränken sie sich nicht auf Ressourcen, die im Web abrufbar sind (z.B. Webseiten), sondern können auch abstrakte Ressourcen, die nicht physisch existieren (z.B. das Konzept "Autor") benennen.

2.3 Resource Description Framework

RDF ist ein Framework zur Repräsentation von Information im World Wide Web. Der Vorteil der Verwendung von RDF zur Bereitstellung von Daten und Metadaten liegt darin, dass Programme, wie zum Beispiel Softwareagenten, diese verarbeiten und ohne Bedeutungsverlust austauschen können. Dadurch wird der Wert dieser Informationen gesteigert und die Funktionalität sowie Interoperabilität des World Wide Web verbessert.²¹

21 RDF Primer: <http://www.w3.org/TR/rdf-primer>

2.3.1 Datenmodell

RDF basiert auf einem graphbasierten Datenmodell.²² Ein RDF Graph ist eine Menge von Tripeln. Ein Tripel besteht aus Subjekt, Prädikat und Objekt. Subjekt und Objekt sind Knoten, das Prädikat ist eine zum Objekt gerichtete Kante, die eine Relation zwischen Subjekt und Objekt zum Ausdruck bringt. Solche expliziten Assoziationen zwischen ENTITÄTEN können beispielsweise in XML nicht ausgedrückt werden. Eine Menge von Tripeln ist damit ein gerichteter und beschrifteter RDF Graph. Die Aussage eines RDF Graphen ist die Konjunktion aller Tripel, die er enthält.

Zur Veranschaulichung soll ein kleines Beispiel dienen:

Diese Arbeit, die durch den URI <http://www.stud.fernuni-hagen.de/q6261035/Bachelorarbeit/Nachrichtenvermittlung> identifiziert wird, hat die Autorin Kathrin Dentler, die durch ihre E-Mail-Adresse kathrin@nulllogic.de eindeutig bezeichnet ist.

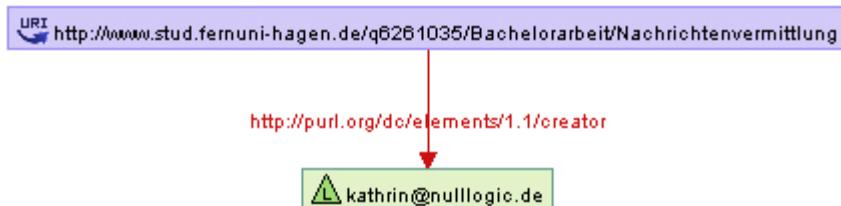


Abbildung 2: Ein einfaches RDF Statement als Graph dargestellt²³

Diese Aussage kann, wie aus Abbildung 2 ersichtlich, als RDF Statement dargestellt werden und besteht aus:

- dem Subjekt <http://www.stud.fernuni-hagen.de/q6261035/Bachelorarbeit/Nachrichtenvermittlung>,
- dem Prädikat <http://purl.org/dc/elements/1.1/creator>, das aus dem DUBLIN CORE stammt
- und dem Objekt kathrin@nulllogic.de, das ein Literal ist.

Eine Ressource ist durch eine URI reference eindeutig identifizierbar. Das ist ein URI mit optionalem, durch das Zeichen "#" getrennten Fragment Identifier. Ein Fragment Identifier verweist auf einen Bereich des Dokuments, das durch den URI bezeichnet wird. RDF URI

22 Resource Description Framework (RDF): Concepts and Abstract Syntax: <http://www.w3.org/TR/rdf-concepts>

23 Die Graphik wurde mit RDF Gravity (RDF Graph Visualization Tool) erstellt: <http://semweb.salzburgresearch.at/apps/rdf-gravity>

references dürfen UNICODE Zeichen beinhalten, wodurch mehrere Sprachen benutzt werden können. Subjekte und Prädikate müssen RDF URI references, Objekte können auch Literale sein. Durch die Verwendung von URI references ist das Vokabular erweiterbar und jeder Anwender kann Aussagen über jede beliebige Ressource treffen.

Es gilt:

- Das Subjekt ist eine RDF URI reference oder ein blank node.
- Das Prädikat ist eine RDF URI reference.
- Das Objekt ist eine RDF URI reference, ein blank node oder ein Literal.

Ein blank node ist ein Knoten ohne URI reference, der innerhalb eines Graphs mit einem blank node identifier mit der Form `_:name` eindeutig bezeichnet wird. Blank nodes werden verwendet, um mehrstellige Relationen als Tripel auszudrücken.

Literale sind Strings (Zeichenketten), die einfach oder typisiert sein können. An ein typisiertes Literal ist die URI reference eines Datentyps angehängt. Hier werden oft die vordefinierten Datentypen des XML Schemas verwendet. RDF hat nur einen integrierten Datentyp, XMLLiteral, der es erlaubt, XML in Literalen zu verwenden.

Zudem bietet RDF eine Technik, die als Reification (Vergegenständlichung) bezeichnet wird, um Aussagen über Aussagen zu treffen, wie zum Beispiel: die Fernuniversität glaubt, dass `kathrin@nulllogic.de` der Autor der Bachelorarbeit Nachrichtenvermittlung ist.

Statements, über die weitere Aussagen getroffen werden sollen, werden dazu als Ressourcen deklariert und können dann als solche verwendet werden. Dies ist besonders interessant für die oberste Schicht des Semantic Web, Trust (siehe Abbildung 1), da hiermit Aussagen von unabhängigen Instanzen bewertet und weiter beschrieben werden können. Wenn Instanzen, denen man vertraut, beispielsweise einen Online Shop als vertrauenswürdig einstufen, darf man je nach Ermessen selbst darauf vertrauen, dort ohne Risiko einkaufen zu können.

2.3.2 RDF/XML

RDF hat eine empfohlene XML basierte Syntax, mit der Graphen serialisiert werden können, um sie zu notieren und auszutauschen. Obiges Beispiel wird in RDF/XML folgenderweise serialisiert:

```

<?xml version="1.0"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/">

  <rdf:Description rdf:about="http://www.stud.fernuni-hagen.de/
    q6261035/Bachelorarbeit/Nachrichtenvermittlung">
    <dc:creator>kathrin@nulllogic.de</dc:creator>
  </rdf:Description>
</rdf:RDF>24
```

Ein RDF/XML Dokument besteht aus einem `rdf:RDF` Element als Wurzel, die `rdf:Description` Elemente enthält. Jede dieser Descriptions trifft Aussagen über ein Subjekt, das im `rdf:about` Attribut identifiziert ist.

Es sei bemerkt, dass RDF/XML nicht die einzige Möglichkeit ist, RDF zu serialisieren. Eine einfache Alternative bietet die von Tim Berners-Lee entworfene Notation 3:

```

@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .

<http://www.stud.fernuni-hagen.de/q6261035/
  Bachelorarbeit/Nachrichtenvermittlung>
dc:creator "kathrin@nulllogic.de" .25
```

2.3.3 RDF Schema

RDF Schema, die RDF Vocabulary Description Language²⁶, basiert auf RDF und bietet die Möglichkeit, domänen spezifisches Vokabular zu definieren, das in RDF Datenmodellen verwendet wird und deren Interpretation dient.

Mit RDF Schema lassen sich Taxonomien und einfache Ontologien formalisieren. RDF Schema ist eine semantische Erweiterung von RDF, mit der semantische Information für Anwendungen zugänglich gemacht wird.

Eine der bekanntesten Anwendungen von RDF Schema ist RSS 1.0²⁷ (RDF Site Summary). Viele Menschen besuchen täglich viele Seiten im Web, um sich über Neuigkeiten zu

24 Dieses Beispiel ist unter <http://www.stud.fernuni-hagen.de/q6261035/Bachelorarbeit/beispiel.rdf> zu finden und wurde mit dem W3C RDF Validation Service validiert: <http://www.w3.org/RDF/Validator>

25 Dieses Beispiel liegt unter <http://www.stud.fernuni-hagen.de/q6261035/Bachelorarbeit/beispieln3.txt> und wurde mit dem Mindswap RDF Converter konvertiert: <http://www.mindswap.org/2002/rdfconvert>

26 RDF Vocabulary Description Language 1.0: <http://www.w3.org/TR/rdf-schema>

27 RSS 1.0: <http://web.resource.org/rss/1.0/spec>

informieren. RSS ist ein RDF Vokabular, um Inhalte zur Verteilung und Wiederverwendung anzubieten. Diese Inhalte werden dabei als so genannte Newsfeeds zur Verfügung gestellt, so dass sie von unabhängigen Programmen, wie zum Beispiel Webseiten, die aktuelle Neuigkeiten zusammenstellen, verwendet werden können. Zudem gibt es immer mehr eigenständige Programme (Feed Reader), mit denen die Anwender Feeds abonnieren und damit eine Übersicht der für sie interessanten Seiten erlangen können, ohne jede dieser Seiten einzeln aufrufen zu müssen.

Eine weitere bekannte Anwendung ist FOAF²⁸ (Friend of a Friend), mit der ein Netz maschinenlesbarer Dokumente über Menschen und ihre Beziehungen geschaffen wird. Kernstück von FOAF ist das FOAF Vokabular, eine in RDF Schema geschriebene Ontologie. Mit diesem Vokabular kann man sich selbst und seine Freunde beschreiben, indem man auf deren FOAF Dokumente verweist. Zudem gibt es verschiedene Tools, um FOAF Dokumente im RDF Format zu erstellen, sie anzuzeigen und zu traversieren.

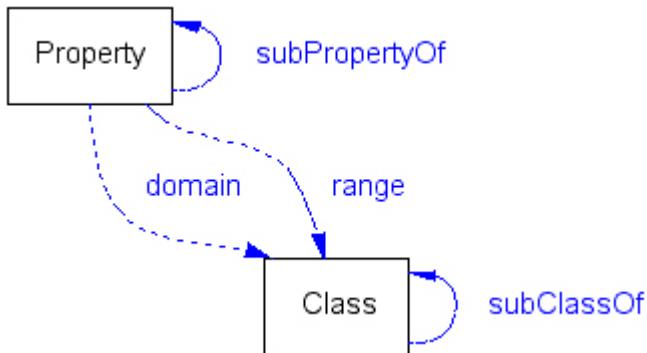


Abbildung 3: RDF Schema²⁹

Die grundlegenden Konzepte von RDFS sind, wie in Abbildung 3 dargestellt, Klassen und Eigenschaften sowie Klassen- und Eigenschaftshierarchien, wobei auch Mehrfachvererbung unterstützt wird. Mit `rdfs:domain` wird die Klasse des Anwendungsbereichs einer Eigenschaft festgelegt. Mit `rdfs:range` wird die Klasse des Wertebereichs einer Eigenschaft festgelegt.

28 FOAF: <http://www.foaf-project.org>

29 Die Graphik wurde mit dem FRODO RDFSviz Tool erstellt: <http://www.dfki.uni-kl.de/frodo/RDFSviz>

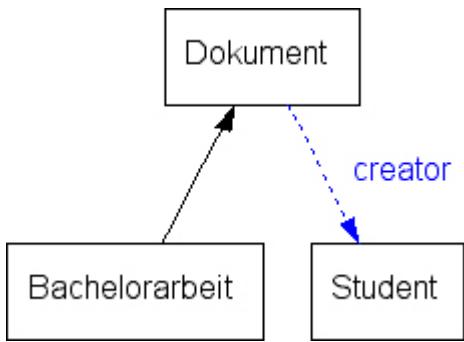


Abbildung 4: Ein einfaches RDF Schema³⁰

Anhand des obigen Beispiels soll jetzt die Anwendungsdomäne genauer beschrieben werden. Dokument, Student und Bachelorarbeit werden als Klassen deklariert. Zudem wird festgelegt, dass eine Bachelorarbeit eine Subklasse der Klasse Dokument ist (`rdfs:subClassOf`). Damit ist jede Instanz der Klasse Bachelorarbeit auch eine Instanz der Klasse Dokument. Für das Prädikat `creator` aus dem Dublin Core wird festgelegt, dass es als Subjekt (`rdfs:domain`) ein Dokument und als Objekt (`rdfs:range`) einen Studenten haben muss. Dies zeigt die Erweiterbarkeit von RDF: ein Prädikat aus einem fremden Vokabular kann für eigene Zwecke individuell angepasst werden. Abbildung 4 zeigt das zugehörige RDF Schema.

Dieses wird in RDF/XML folgenderweise serialisiert:

```

<?xml version="1.0"?>
<rdf:RDF
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
    xml:base="http://www.stud.fernuni-hagen.de/
    q6261035/Bachelorarbeit/stud.rdfs#">

    <rdfs:Class rdf:ID="Dokument"></rdfs:Class>
    <rdfs:Class rdf:ID="Student"></rdfs:Class>

    <rdfs:Class rdf:ID="Bachelorarbeit">
        <rdfs:subClassOf rdf:resource="Dokument"/>
    </rdfs:Class>

    <rdf:Property rdf:about="http://purl.org/dc/elements/1.1/creator">
        <rdfs:domain rdf:resource="Dokument"/>
        <rdfs:range rdf:resource="Student"/>
    </rdf:Property>
</rdf:RDF>31

```

30 Die Graphik wurde mit dem FRODO RDFSviz Tool erstellt: <http://www.dfki.uni-kl.de/frodo/RDFSviz>

31 Dieses Beispiel liegt unter <http://www.stud.fernuni-hagen.de/q6261035/Bachelorarbeit/stud.rdfs> und wurde mit dem W3C RDF Validation Service validiert: <http://www.w3.org/RDF/Validator>

Jetzt wird in dem gleichen Beispiel das RDF Schema angewandt und es werden Instanzen aus Klassen des Schemas gebildet:

```
<?xml version="1.0"?>
<rdf:RDF
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:stud="http://www.stud.fernuni-hagen.de/
        q6261035/Bachelorarbeit/stud.rdfs#"
    xmlns:dc="http://purl.org/dc/elements/1.1/"
    xml:base="http://www.stud.fernuni-hagen.de/q6261035/Bachelorarbeit/">

    <stud:Bachelorarbeit rdf:about="Nachrichtenvermittlung">
        <dc:creator>
            <stud:Student rdf:about="kathrin@nulllogic.de"/>
        </dc:creator>
    </stud:Bachelorarbeit>
</rdf:RDF>32
```

"Nachrichtenvermittlung" ist nun eine Instanz der Klasse Bachelorarbeit und damit wegen der Subklassenbeziehung auch ein Dokument, dessen Autor konform mit dem verwendeten Schema ein Student ist.

Die Angabe von Domain und Range im RDF Schema kann dazu dienen, Typen von Instanzen zu erschließen. Aus dem obigen RDF Dokument kann man beispielsweise folgern, dass die Ressource <http://www.stud.fernuni-hagen.de/q6261035/Bachelorarbeit/Nachrichtenvermittlung> vom Typ Dokument ist, da als Domain der Eigenschaft creator die Klasse Dokument angegeben wurde.

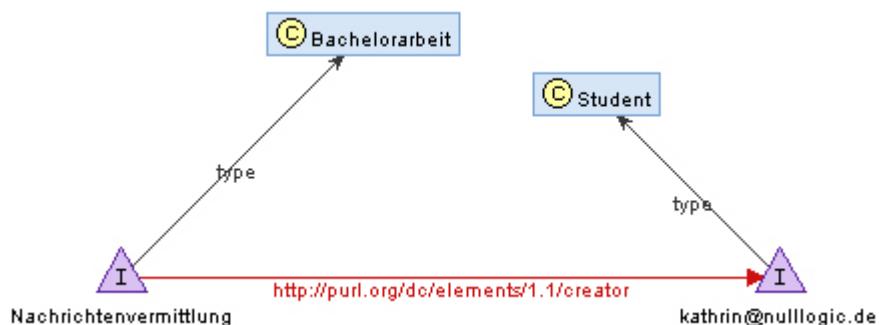


Abbildung 5: Verwendung von RDF Schema³³

32 Das RDF Dokument liegt unter <http://www.stud.fernuni-hagen.de/q6261035/Bachelorarbeit/stud.rdf> und wurde mit dem ICS FORTH VRP validiert: <http://139.91.183.30:9090/RDF/VRP>

33 Die Graphik wurde mit RDF Gravity (RDF Graph Visualization Tool) erstellt: <http://semweb.salzburgresearch.at/apps/rdf-gravity>

2.3.4 Semantik

RDF/S hat eine formale Semantik und damit ein komplettes System von Inferenzregeln.³⁴

Diese formale Semantik wird mit Hilfe der Modelltheorie spezifiziert. Die Modelltheorie nimmt an, dass sich eine Sprache auf eine "Welt" bezieht und beschreibt die Mindestanforderungen, die eine Welt erfüllen muss, um jedem Ausdruck der Sprache eine angemessene Bedeutung zuzuweisen. Eine bestimmte Welt heißt Interpretation. Eine Aussage läuft auf die Behauptung hinaus, dass die Welt eine Interpretation ist, die die Aussage wahr sein lässt. Die Idee ist es, eine abstrakte, mathematische Beschreibung der Eigenschaften zur Verfügung zu stellen, die eine solche Interpretation haben muss. Damit lässt sich bestimmen, wann INFERENZEN gültig sind.

Eine einfache Interpretation I eines Vokabulars V ist folgenderweise definiert (einfache Literale mit language tags, typisierte Literale und Graphen mit blank nodes werden bewusst ausgelassen):

- 1 IR: eine nicht leere Menge von Ressourcen, auch Domäne oder Universum von I genannt
- 2 IP: die Menge der Eigenschaften von I
- 3 IEXT: Abbildung von IP in die Potenzmenge von $IR * IR$
- 4 IS: Abbildung der URI references aus V in $(IR \cup IP)$
- 5 LV: Teilmenge aus IR, die alle einfachen Literale aus V enthält

$EXT(x)$, die EXTENSION von x , ist die Menge aller Paare derjenigen Argumente, für die die Eigenschaft wahr ist.

Eine Interpretation stellt genügend Information über eine mögliche Welt zur Verfügung, um jedem Tripel eines RDF Graphen und damit dem gesamten Graphen einen Wahrheitswert zuzuweisen. Die Bedeutung eines Graphen wird rekursiv aus seinen syntaktischen Bestandteilen gewonnen. Für Graphen gelten folgende semantische Bedingungen, wobei E für RDF Syntax steht:

- Wenn E ein einfaches Literal "aaa" aus V ist, dann gilt $I(E) = aaa$
- Wenn E eine URI reference aus V ist, dann gilt $I(E) = IS(E)$
- Wenn E ein Tripel (S, P, O) ist, dann gilt $I(E) = \text{wahr}$, wenn S, P und O aus V sind, $I(P)$ aus IP und das geordnete Paar $\langle I(S), I(O) \rangle$ aus $IEXT(I(P))$ ist; ansonsten gilt $I(E) = \text{falsch}$

34 RDF Semantics: <http://www.w3.org/TR/rdf-mt>

- Wenn E ein Graph ist, dann gilt $I(E) = \text{wahr}$, wenn $I(E') = \text{wahr}$ für alle Tripel E' in E gilt; ansonsten gilt $I(E) = \text{falsch}$

I erfüllt E , wenn $I(E) = \text{wahr}$, und aus einer Menge S von RDF Graphen folgt logisch der Graph E , wenn jede Interpretation, die jeden Graphen aus S erfüllt, auch E erfüllt. Damit ergeben sich einige einfache Ergebnisse über das logische Folgern und die gültige Inferenz. Das Hauptergebnis ist das Interpolation Lemma:

Aus S folgt logisch genau dann der Graph E , wenn ein Subgraph von S eine Instanz von E ist.

Das RDF Vokabular ist die Menge der URI references aus dem `rdf` Namespace. Eine RDF Interpretation ist eine einfache Interpretation, die zusätzliche semantische Bedingungen erfüllt. Semantische Bedingungen, die sich auf `rdf:XMLLiteral` beziehen, werden bewusst ausgelassen. Somit ist die verbleibende semantische Bedingung:

- x ist in IP genau dann, wenn $\langle x, I(\text{rdf:Property}) \rangle$ in $\text{IEXT}(I(\text{rdf:type}))$ ist.

Auch das RDFS Vokabular fügt den obigen semantischen Bedingungen weitere hinzu. Neu ist IC , die Menge aller Klassen einer Interpretation. Literale und Datentypen werden auch hier bewusst ausgelassen:

- x ist in $\text{ICEXT}(y)$ genau dann, wenn $\langle x, y \rangle$ in $\text{IEXT}(I(\text{rdf:type}))$ ist
- $IC = \text{ICEXT}(I(\text{rdfs:Class}))$
- $R = \text{ICEXT}(I(\text{rdfs:Resource}))$
- Wenn $\langle x, y \rangle$ in $\text{IEXT}(I(\text{rdfs:domain}))$ und $\langle u, v \rangle$ in $\text{IEXT}(x)$ sind, dann ist u in $\text{ICEXT}(y)$
- Wenn $\langle x, y \rangle$ in $\text{IEXT}(I(\text{rdfs:range}))$ und $\langle u, v \rangle$ in $\text{IEXT}(x)$ sind, dann ist v in $\text{ICEXT}(y)$
- $\text{EXT}(I(\text{rdfs:subPropertyOf}))$ ist transitiv und reflexiv bezüglich IP
- Wenn $\langle x, y \rangle$ in $\text{IEXT}(I(\text{rdfs:subPropertyOf}))$ ist, dann sind x und y in IP und $\text{ICEXT}(x)$ ist eine Teilmenge von $\text{IEXT}(y)$
- Wenn x in IC ist, dann ist $\langle x, I(\text{rdfs:Resource}) \rangle$ in $\text{IEXT}(I(\text{rdfs:subClassOf}))$
- Wenn $\langle x, y \rangle$ in $\text{IEXT}(I(\text{rdfs:subClassOf}))$ ist, dann sind x und y in IC und $\text{ICEXT}(x)$ ist eine Teilmenge von $\text{ICEXT}(y)$
- $\text{IEXT}(I(\text{rdfs:subClassOf}))$ ist transitiv und reflexiv bezüglich IC

Zudem gibt es für das RDF und RDFS Vokabular AXIOMATISCHE Tripel, die in der Interpretation erfüllt sein müssen, zum Beispiel:

```
rdfs:domain rdfs:range rdfs:Class .
```

Damit ergeben sich einfache Regeln der Form:

"Füge einem Graph ein Tripel hinzu, wenn er Tripel enthält, die einem bestimmten Muster entsprechen."

Wenn E beispielsweise die Tripel `uuu rdfs:subPropertyOf vvv und vvv rdfs:subPropertyOf www` enthält, dann füge das Tripel `uuu rdfs:subPropertyOf www` hinzu.

2.3.5 SPARQL Query Language

Das Semantic Web ist ein Web aus Daten, ähnlich einer global verteilten Datenbank. SPARQL³⁵ als Abfragesprache für RDF Daten bietet die Möglichkeit, mit Abfragen auf der semantischen Ebene Information aus dem Web zu ermitteln und zu kombinieren.

Im Gegensatz zu XML, RDF und OWL ist SPARQL noch keine Recommendation des W3C, sondern ein Working Draft, also ein Arbeitsentwurf, der im W3C die erste Vorstufe im Entwicklungsprozess zur Recommendation darstellt. Deshalb sollen die folgenden Informationen nur als Grundidee für eine RDF Abfragesprache, die für das Semantic Web von essentieller Bedeutung ist, verstanden werden.

Mit SPARQL kann man Informationen aus RDF Graphen abfragen. Es ist möglich, neue Graphen zu konstruieren, die auf Informationen des abgefragten Graphen basieren. Die abzufragenden Tripel können dabei aus mehreren Quellen stammen: direkt aus einem RDF Dokument, abgeleitet von anderen Tripeln oder extrahiert aus weiteren Formaten.

Mit SPARQL lassen sich auch entfernte Abfragen durchführen, indem Web Services genutzt werden. Dabei werden Abfragen zu einem Abfragen verarbeitenden Service geschickt, der seine Ergebnisse zurückgibt.³⁶

So lässt sich beispielsweise der Autor der Bachelorarbeit aus dem Graphen des obigen Beispiels³⁷ abfragen. Die SELECT Klausel bezeichnet die Variablen, die in den Ergebnissen vorkommen sollen, und die WHERE Klausel beinhaltet ein Tripel als Muster:

35 SPARQL Query Language for RDF: <http://www.w3.org/TR/rdf-sparql-query>

36 SPARQL Protocol for RDF: <http://www.w3.org/TR/rdf-sparql-protocol>

37 <http://www.stud.fernuni-hagen.de/q6261035/Bachelorarbeit/beispiel.rdf>

```

SELECT ?creator WHERE {
  <http://www.stud.fernuni-hagen.de/
    q6261035/Bachelorarbeit/Nachrichtenvermittlung>
  <http://purl.org/dc/elements/1.1/creator> ?creator .
}

```

Das Ergebnis³⁸ dieser Abfrage ist "kathrin@nulllogic.de".

2.4 Web Ontology Language

OWL³⁹ wurde aus DAML+OIL (DARPA Agent Markup Language + Ontology Inference Layer) heraus entwickelt und ist die ausdrucksstärkste Ontologiesprache. OWL bietet die Möglichkeit, die Semantik von Termen eines Vokabulars und die Beziehungen dieser Terme explizit zu repräsentieren. Eine solche Repräsentation ist eine Ontologie, mit der die Bedeutung der Terminologie, die in Webdokumenten verwendet wird, formal beschrieben werden kann.

Das Akronym der Web Ontology Language wäre eigentlich WOL. Die Web Ontology Language wird jedoch mit OWL abgekürzt. Dies geschieht in Anlehnung an die Eule (OWL) in Winnie the Pooh, die als einzige im Wald ihren Namen buchstabieren konnte, allerdings mit einem Dreher: WOL. Die Web Ontology Language macht diesen Dreher genau anders herum: OWL.⁴⁰

OWL wurde für Anwendungen konzipiert, die den Inhalt von Information verarbeiten. OWL unterstützt durch zusätzliches Vokabular mit formaler Semantik eine stärkere maschinelle Interpretationsmöglichkeit von Inhalten im Web als die bisher vorgestellten Sprachen.

Eine konkrete Anwendung von OWL ist OWL Service⁴¹, eine Ontologie zur Beschreibung der Eigenschaften und Fähigkeiten von Web Services in einer Form, die von Softwareagenten verarbeitet werden kann. OWL-S macht Web Services zu Semantic Web Services und ermöglicht die Automatisierung des Auffindens, der Ausführung, Komposition und Interoperation von Web Services.

Hier wird zunächst auf Taxonomien und Ontologien eingegangen, die man mit Hilfe von OWL repräsentieren kann.

38 Redland Rasqal RDF Query Demonstration: <http://librdf.org/query>

39 Web Ontology Language (OWL): <http://www.w3.org/2004/OWL>

40 Frequently Asked Questions on W3C's Web Ontology Language (OWL):
<http://www.w3.org/2003/08/owlfaq.html.en>

41 OWL-S: <http://www.daml.org/services/owl-s/1.1>

2.4.1 Taxonomien

In der Biologie ist eine Taxonomie die Einordnung der Lebewesen in systematische Einheiten auf Grund ihrer verwandtschaftlichen Beziehungen [DGKM71].

In der Informatik spricht man von einer Taxonomie als "Klassifikation von Entitäten in Form einer Hierarchie" [DOS03]. Damit können Subklassifizierungen und Subklassen ausgedrückt und die entsprechenden Entitäten mit Semantik und Struktur versehen werden. Taxonomien sind das Grundgerüst von Ontologien.

Taxonomien werden vor allem zur Ablage und Suche verwendet, wie zum Beispiel in Bibliotheken oder in dem von Netscape initiierten Open Directory Project, dessen Ziel die Erstellung des umfangreichsten Internet-Verzeichnisses mit Hilfe einer großen Gemeinschaft freiwilliger Editoren ist.⁴²

2.4.2 Tagging

Tagging wird im Englischen gerne mit dem Neologismus Folksonomy⁴³ (Folk und Taxonomy) bezeichnet. Es handelt sich hierbei um einen gemeinschaftlichen und einfachen Mechanismus, um Informationen im Web zu kategorisieren. Beim Tagging können Nutzer ihrer Information frei gewählte Schlagworte, so genannte Tags, zuordnen. Tagging wird in Anwendungen genutzt, in denen Bilder, Bookmarks oder Artikel publiziert werden. Auch in sozialen Anwendungen, die menschliche Kommunikation, Interaktion und Zusammenarbeit unterstützen, spielt Tagging eine immer wichtigere Rolle.

Der Nachteil von Tagging im Vergleich zu kontrollierten Vokabularen liegt in der fehlenden Struktur der verwendeten Schlagworte, zwischen denen sich keine Beziehungen wie beispielsweise Subklassen darstellen lassen.

Dem gegenüber stehen, wie die Praxis zeigt, mehrere Vorteile. So ist Tagging einfach zu verstehen und zu verwenden. Aus modernen Anwendungen, in denen Tagging genutzt wird, ist ersichtlich, dass die Akzeptanz der Nutzer groß ist. Beispiele sind Flickr⁴⁴, eine Webanwendung zur Veröffentlichung von digitalen Bildern, oder Lycos iQ, ein Fragen- und Antwortensystem. Die Kategorien beim Tagging sind nicht wie in einer Taxonomie vorgegeben, sondern entwickeln sich im Laufe der Zeit demokratisch und dynamisch.

42 Open Directory Project: <http://dmoz.org/about.html>

43 Wikipedia Folksonomy: <http://de.wikipedia.org/wiki/Folksonomy>

44 Flickr: <http://www.flickr.com>

2.4.3 Ontologien

Ontologie ist eine philosophische Disziplin, die sich mit dem Sein, dem Seienden als solchem und mit den fundamentalen Typen von Entitäten beschäftigt.⁴⁵

Dieser Begriff wurde von der Informatik übernommen und mit einer spezifischen technischen Bedeutung versehen. Statt von "Ontologie" spricht man hier von "einer Ontologie". Eine beliebte Definition ist die von T. R. Gruber⁴⁶, überarbeitet von R. Studer [AH04]:

An ontology is an explicit and formal specification of a conceptualization.

Eine Ontologie ist demnach eine explizite formale Beschreibung von Konzepten einer bestimmten Anwendungsdomäne und Beziehungen zwischen diesen Konzepten. Damit ist eine Ontologie ein abstraktes Modell einer Domäne. Heterogene und verteilte Programme, die dieselbe Ontologie verwenden, haben ein gemeinsames Vokabular sowie ein gemeinsames "Verständnis" von dieser Anwendungsdomäne, so dass sie sich gegenseitig Abfragen stellen und Aussagen treffen können, die mit der Ontologie konsistent sind.

Ontologien stellen im Semantic Web die konzeptuelle Basis dar, die es Maschinen ermöglicht, die Semantik von Metadaten zu interpretieren [SS04].

Verschiedene Terminologien können auf eine Ontologie und verschiedene Ontologien aufeinander abgebildet werden. Dies ermöglicht semantische Interoperabilität.

2.4.4 OWL Vokabular

RDF Schema ist ein Vokabular, um Klassen und Eigenschaften sowie Klassen- und Eigenschaftshierarchien zu definieren und damit einfache Ontologien zu erstellen. OWL bietet darüber hinaus weiteres Vokabular, um Klassen und Eigenschaften zu beschreiben: unter anderem Beziehungen zwischen Klassen (z.B. DISJUNKTHEIT), KARDINALITÄT ("genau eins"), Äquivalenz oder Charakteristika von Eigenschaften (z.B. Transitivität und Symmetrie).⁴⁷

Eigenschaften

In OWL gibt es zwei Arten von Eigenschaften: Objekteigenschaften, die Objekte miteinander in Beziehung setzen, und Datentypeigenschaften, die Objekten Datentypen zuweisen. Eigenschaften können mehrere Domains oder Ranges haben, wobei deren Schnittmenge gebildet wird.

45 Wikipedia Ontologie: <http://de.wikipedia.org/wiki/Ontologie>

46 What is an Ontology: <http://www.ksl.stanford.edu/kst/what-is-an-ontology.html>

47 OWL Web Ontology Language Guide: <http://www.w3.org/TR/owl-guide>

Für die weitere Spezifikation der Charakteristika von Eigenschaften stellt OWL folgende Mechanismen zur Verfügung:

- TransitiveProperty

Wenn eine Eigenschaft P transitiv ist, dann gilt für jedes x, y und z:

$P(x, y) \text{ und } P(y, z) \text{ impliziert } P(x, z)$

- SymmetricProperty

Wenn eine Eigenschaft P symmetrisch ist, dann gilt für jedes x und y:

$P(x, y) \text{ genau dann, wenn } P(y, x)$

- FunctionalProperty

Wenn eine Eigenschaft P funktional ist, dann gilt für alle x, y und z:

$P(x, y) \text{ und } P(x, z) \text{ impliziert } y = z$

- inverseOf

Ist eine Eigenschaft P1 invers zu einer Eigenschaft P2, dann gilt für alle x und y:

$P1(x, y) \text{ genau dann, wenn } P2(y, x)$

- InverseFunctionalProperty

Ist eine Eigenschaft P invers funktional, dann gilt für alle x, y und z:

$P(y, x) \text{ und } P(z, x) \text{ impliziert } y = z$

Restriktionen

Restriktionen definieren anonyme Klassen (blank nodes) und schränken im Kontext einer `owl:Restriction` die Verwendung von Eigenschaften (`owl:onProperty`) für diese Klassen in Bezug auf deren Range ein.

Die Restriktion `owl:allValuesFrom` legt den Typ des Wertes einer Eigenschaft in Verbindung mit einer bestimmten Klasse fest. Dies entspricht einer universellen Quantifikation. Analog dazu wird mit der Restriktion `owl:someValuesFrom` für eine Eigenschaft in Verbindung mit einer Klasse festgelegt, dass mindestens ein Wert der betroffenen Eigenschaft den angegebenen Typ hat. Dies entspricht einer existentiellen Quantifikation.

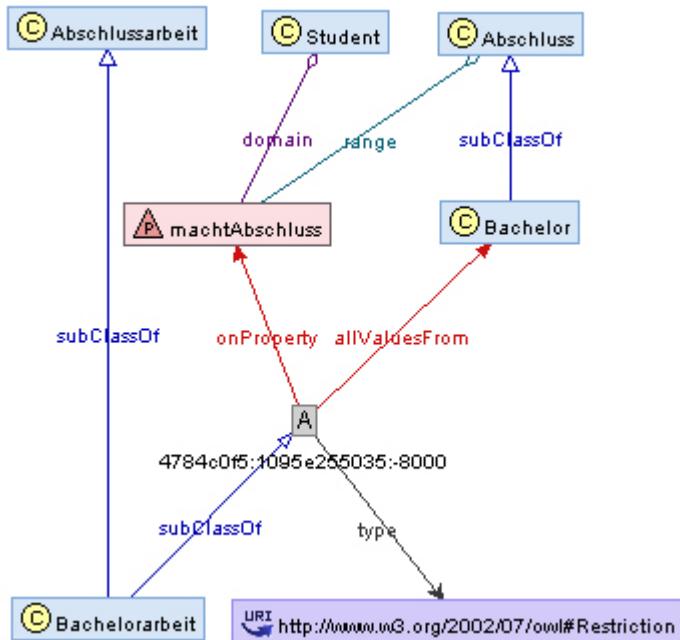


Abbildung 6: Eine einfache Restriktion⁴⁸

Mit einer Restriktion kann beispielsweise definiert werden, dass eine Bachelorarbeit Subklasse der Klasse Abschlussarbeit ist und von einem Studenten geschrieben werden muss, der als Abschluss den Bachelor anstrebt. Dieser Graph wird in OWL/XML folgenderweise serialisiert:

```

<owl:Class rdf:id="Bachelorarbeit">
  <rdfs:subClassOf>
    <owl:Class rdf:id="Abschlussarbeit"/>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty>
        <owl:ObjectProperty rdf:id="machtAbschluss"/>
      </owl:onProperty>
      <owl:allValuesFrom rdf:resource="#Bachelor"/>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>49

```

Mit demselben Mechanismus werden auch Kardinalitäten definiert.

48 Diese Graphik wurde mit RDF Gravity (RDF Graph Visualization Tool) erstellt:
<http://semweb.salzburgresearch.at/apps/rdf-gravity>

49 Die gesamte Ontologie ist unter <http://www.stud.fernuni-hagen.de/q6261035/Bachelorarbeit/university.owl> zu finden, wurde mit Protégé erstellt:
<http://protege.stanford.edu> und mit dem WonderWeb OWL Validator validiert:
<http://phoebe.cs.man.ac.uk:9999/OWL/Validator>

Im obigen RDF Schema Beispiel wurde für die Eigenschaft creator aus dem Dublin Core der Wertebereich (Range) Student festgelegt. Dies ist eine globale Restriktion, da sie überall da gilt, wo das Schema verwendet wird. OWL Restriktionen hingegen sind lokal: für die Klasse Bachelorarbeit muss der Autor ein Student sein, der als Abschluss den Bachelor macht. Für die Klasse Masterarbeit kann festgelegt werden, dass der Range der Eigenschaft machtAbschluss die Klasse Master ist.

Komplexe Klassen

OWL stellt Konstruktoren zur Verfügung, um so genannte Klassenausdrücke zu erstellen. OWL unterstützt die wesentlichen Mengenoperationen Vereinigung, Schnitt und Komplement. Auch die Disjunktion zwischen den Extensionen von Klassen kann ausgedrückt werden.

Mapping

Um Ontologien bestmöglich zu nutzen, müssen sie gemeinsam verwendet, wieder verwendet und zusammengesetzt werden können. Dazu bietet OWL eine Reihe von Möglichkeiten: so können Klassen und Eigenschaften als äquivalent und Individuen als identisch oder verschieden deklariert werden, um unterschiedliche Terminologien aus verschiedenen Ontologien aufeinander abzubilden und diese Ontologien damit zu vereinen. OWL geht nicht von eindeutigen Namen aus, was bedeutet, dass zwei unterschiedliche Namen nicht notwendigerweise zwei unterschiedliche Individuen bezeichnen.

In ein OWL Dokument können weitere OWL Ontologien importiert und damit verwendet werden. OWL geht von der Open-World-Assumption aus: Beschreibungen von Ressourcen beschränken sich nicht auf ein spezielles Dokument oder einen Geltungsbereich. OWL ist monoton, das heißt, dass Aussagen oder Schlussfolgerungen nur hinzukommen, nicht aber rückgängig gemacht werden können.

2.4.5 Untersprachen von OWL

Bei Ontologiesprachen gibt es einen Trade-off (negative wechselseitige Abhängigkeit zweier Aspekte) zwischen Ausdrucksstärke und effizientem Folgern. Je mächtiger eine Sprache ist, desto weniger effizient können Schlussfolgerungen gezogen werden, wobei es möglich ist, die Grenzen der Berechenbarkeit zu überschreiten. Diesen Trade-off hat das W3C mit der Definition mehrerer Untersprachen mit zunehmender Ausdrucksstärke gelöst:

- OWL Lite

OWL Lite beschränkt sich auf eine Teilmenge der OWL Konstruktoren, was die Sprache auf der einen Seite einfach zu handhaben und zu implementieren macht, auf der anderen Seite jedoch in ihrer Ausdrucksstärke beschneidet.

- OWL DL
OWL Description Logic (entscheidbare Teilmenge der Prädikatenlogik) nutzt alle Elemente des OWL Vokabulars, jedoch in einer eingeschränkten Form. So wird beispielsweise die Anwendung von OWL Konstruktoren aufeinander nicht ermöglicht (eine Klasse kann nicht Instanz einer anderen Klasse sein). Dadurch wird effizientes Schlussfolgern gewährleistet, was bedeutet, dass jede gültige Folgerung in endlicher Zeit durchgeführt werden kann.
- OWL Full
OWL Full nutzt uneingeschränkt alle Elemente des OWL Vokabulars. Diese können mit RDF und RDFS kombiniert und aufeinander angewandt werden. OWL Full ist weder entscheidbar noch komplett implementierbar.

Jede dieser Untersprachen ist in Bezug auf korrekte Ausdrücke und gültige Schlussfolgerungen eine Erweiterung des einfacheren Vorgängers.

OWL Full kann als Erweiterung von RDF betrachtet werden, während dies für OWL Lite und OWL DL nur eingeschränkt gilt. Jedes OWL Dokument ist ein RDF Dokument und jedes RDF Dokument ist ein OWL Full Dokument, jedoch nicht unbedingt ein gültiges OWL Lite oder OWL DL Dokument. Dies liegt an den Möglichkeiten zur METAMODELLIERUNG von RDFS.

2.5 Logik, Regeln und Beweise

Logik, vor allem die Prädikatenlogik, ist eine wichtige Grundlage der Wissensrepräsentation. RDF und OWL (Lite und DL) können als Spezialisierungen der Prädikatenlogik betrachtet werden. Logik stellt formale Sprachen mit formaler Semantik zur Verfügung, um Wissen auszudrücken und aus diesem Wissen Schlussfolgerungen zu ziehen. So kann implizites Wissen explizit gemacht und damit neues Wissen erschlossen werden.

Dies ist besonders interessant für Agenten, die durch Ontologien Wissen über den Inhalt von Ressourcen im Web und über die Konzepte von Anwendungsdomänen und deren Beziehungen gewinnen können, das über das hinausgeht, was tatsächlich in der Ontologie beschrieben ist.

Ein wichtiger Punkt, der sich noch im Aufbau befindet, ist die Verwendung von Regeln. Agenten können nach festen Regeln Entscheidungen treffen und diese dem Benutzer gegenüber begründen, da sich die einzelnen Schritte, die zu der Entscheidung geführt haben, zurückverfolgen lassen. Auch für Interaktionen zwischen Agenten sind Begründungen und Beweise wichtig. So kann es für einen Agenten notwendig sein, zu validieren, ob der Schluss, zu dem ein anderer Agent gekommen ist, richtig ist. Dafür müssen Agenten ihre Schlüsse beweisen und diese Beweise derart repräsentieren, dass andere Agenten diese nachvollziehen können.

Eine einfache Regel hat die Form:

$A \rightarrow B$,

wobei A die Prämisse und B die Konklusion ist. Diese wird als deduktive Regel folgendermaßen interpretiert: wenn A wahr ist, ist auch B wahr.

Eine Teilmenge der Prädikatenlogik ist die Hornlogik. Eine Hornformel ist eine Formel in konjunktiver Normalform (Konjunktion von Disjunktionen), in der jedes Disjunktionsglied eine Hornklausel ist, also eine Klausel mit höchstens einem positiven Literal. Eine Hornklausel kann als Regel betrachtet werden, in der die negativen Literale die Prämisse und das positive Literal die Konklusion darstellen. Regeln mit leerer Prämissenmenge sind Fakten. Diese Regeln und Fakten ergeben zusammen eine Wissensbasis, an die man Abfragen stellen kann.

Aus Fakten in der Form

`mutter(X, Y)`

kann man anhand von Regeln auf verwandtschaftliche Beziehungen schließen. So ist zum Beispiel eine Mutter ein Elternteil:

`mutter(X, Y) → elternteil(X, Y)`

und eine Großmutter die Mutter eines Elternteils:

`mutter(X, E), elternteil(E, Y) → großmutter(X, Y)`

Die Rule Markup Language RuleML⁵⁰ ist eine Auszeichnungssprache für Regeln und dient deren Austausch zwischen verschiedenen Anwendungen.

Die Semantic Web Rule Language SWRL⁵¹ ist eine W3C Member Submission und daher lediglich als Vorschlag für eine mögliche Regelsprache zu betrachten. SWRL basiert auf der Kombination von OWL Lite und OWL DL mit DATALOG RuleML, einer entscheidbaren Untersprache der Rule Markup Language. SWRL ermöglicht die Kombination hornähnlicher Regeln mit einer OWL Wissensbasis. Dadurch wird die Ausdrucksstärke von OWL erhöht, da OWL alleine keine Möglichkeit zur Verfügung stellt, Regeln zu formulieren. Eine Regel in SWRL ist eine Implikation zwischen den Prämissen und der Konklusion, die aus Atomen

50 RuleML: <http://www.ruleml.org>

51 A Semantic Web Rule Language: <http://www.w3.org/Submission/SWRL>

bestehen. Atome können OWL Descriptions, Eigenschaften, Individuen oder Datenwerte sein. Eine beispielhafte Regel besagt, dass der Bruder eines Elternteils ein Onkel ist:

```
elternteil(X, Y), bruder(Z, X) -> onkel(Z, Y)
```

Die Prädikatenlogik ist monoton, das heißt, dass bisherige Folgerungen immer wahr bleiben. Auch nichtmonotone Regeln sind hilfreich, wenn verfügbare Informationen so wie im Semantic Web nicht komplett sind. Auftretende Konflikte können beispielsweise mit Prioritäten gelöst werden, die sich in RuleML ausdrücken lassen. Ein Beispiel sind die folgenden Regeln:

```
Pinguin(X) -> Vogel(X)
Vogel(X) -> fliegen(X)
Pinguin(X) -> ¬fliegen(X)
```

Daraus lässt sich sowohl folgern, dass ein Pinguin fliegen kann, als auch, dass er nicht fliegen kann. Ein solcher Konflikt kann gelöst werden, indem die dritte Regel als Ausnahme mit einer höheren Priorität als die zweite Regel versehen wird. Ist zunächst nur klar, dass Tux ein Vogel ist, dann kann dieser auch fliegen. Kommt später die Tatsache hinzu, dass er ein Pinguin ist, wird der vorherige Schluss revidiert und Tux kann nicht mehr fliegen.

2.6 Vertrauen

Vertrauen ist ein subjektives Gefühl, das nicht implementiert werden kann. Das Semantic Web soll jedoch wichtige Grundlagen bieten, um vertrauenswürdig zu sein. Menschen vertrauen durch Erfahrung, Empfehlungen und plausible Erklärungen. Erfahrungen muss jeder für sich sammeln und benötigt dazu eventuell einen Vertrauensvorschuss beziehungsweise ein gewisses Grundvertrauen.

Vertrauen ist teilweise übertragbar, weshalb Menschen auch auf Empfehlung vertrauen. Das ist die Grundidee der digitalen Signaturen, mit denen sich ein Vertrauensnetz zwischen Menschen und damit ein "Web of Trust" bilden soll. Das Semantic Web ermöglicht es Benutzern mit Hilfe von digitalen Signaturen, Verantwortung für Beiträge zu übernehmen, die sie im Web veröffentlichen. Mit digitalen Signaturen werden Integrität und Authentizität der signierten Daten sowie die Identität des Urhebers sichergestellt, das heißt, die Daten wurden während einer Übertragung nicht verändert und können sicher dem Urheber zugeordnet werden.

Bei der Erstellung einer digitalen Signatur berechnet der Urheber mit Hilfe einer Hash-Funktion aus seinen Daten einen eindeutigen Hash-Wert fester Länge. Dieser Wert ist eine kryptographische Prüfsumme, die als "Fingerabdruck" der Daten dient. Ein Empfänger erhält die Daten mit der zugehörigen digitalen Signatur. Er wendet dieselbe Hash-Funktion auf

diese Daten an. Erhält er denselben Hash-Wert wie den aus der Signatur, ist sichergestellt, dass die Daten nicht verändert wurden.

Um zu gewährleisten, dass die Daten vom angegebenen Urheber stammen, werden zusätzlich Public Key Verschlüsselungsverfahren eingesetzt. Hier stehen ein geheimer und ein öffentlicher Schlüssel zur Verfügung, die beide dem Urheber gehören. Der geheime Schlüssel dient zum Verschlüsseln. Mit dem öffentlichen Schlüssel kann jeder die Daten entschlüsseln und damit sicherstellen, dass die Daten von der Person stammen, die den geheimen Schlüssel besitzt.

Bei der Erzeugung einer digitalen Signatur wird der Hash-Wert der Daten mit dem geheimen Schlüssel des Urhebers verschlüsselt.

Wie aus Abbildung 1 ersichtlich ist, verlaufen digitale Signaturen orthogonal zu den Schichten des Semantic Web, da sie in jeder der Schichten verwendet werden können und sollen. So lassen sich RDF Aussagen, RDF, RDFS und OWL Dokumente sowie Regeln und Beweise signieren. Dies ist gerade in einem weltweit verteilten System wie dem Semantic Web wichtig, in dem jeder Nutzer beliebige Aussagen treffen kann.

Nun stellt sich natürlich die Frage, ob ein Schlüssel tatsächlich dem angegebenen Besitzer gehört und inwiefern man diesem vertrauen kann. Zur Lösung dieser Frage eignen sich verschiedene Vertrauensstufen wie bei PGP (Pretty Good Privacy)⁵², die auch Misstrauen beinhalten. Damit wird ein "Web of Trust" aufgespannt: man vertraut nicht nur denjenigen, denen man direkt vertraut, sondern auch denjenigen, denen diese (direkt) vertrauen und so weiter.

Tim Berners-Lee wünscht sich für Agenten und Browser im Semantic Web einen 'Oh, yeah' - Button, der bei Zweifel und Unsicherheit angeklickt werden kann, um plausible Erklärungen zu liefern. Ist beispielsweise ein Agent durch Regeln und Inferenz zu einem Schluss gekommen, kann er diesen mittels Rückwärtsdeduktion beweisen und dem Anwender die einzelnen Schritte, die zu dem Schluss geführt haben, verständlich darlegen, so dass dieser den Schluss nachvollziehen kann. Zeigt der Browser Informationen aus dem Web, liefert er als Antwort auf den angeklickten Button den Grad, zu dem man der Information vertrauen kann und wie dieser zustande gekommen ist.

52 PGP: <http://www.pgp.com>

3 Global Brain

I've heard it suggested that the next revolution is the formation of a kind of global mind that results from enough people and enough interconnectedness.
Bruce Eckel

Ein Mensch interagiert von Geburt an mit seiner Umwelt. Er lernt die Bedeutung von Wörtern und bildet daraus erste Dreiwortsätze. Damit eröffnet sich die Möglichkeit des inhaltlichen Austauschs mit anderen Menschen. Ein Kind lernt Lesen und Schreiben, also die Techniken, die es erlauben, Inhalte über Raum und Zeit zu bewahren und zu übertragen. Es lernt, mit abstrakten Konzepten und Zusammenhängen umzugehen und dadurch, über sich selbst und über sein Umfeld zu reflektieren. Durch den Gebrauch all dieser Fähigkeiten bilden sich im Laufe der Entwicklung Muster im Gehirn, die man als Wissen bezeichnen kann.

Menschen kennen in ihrem persönlichen Umfeld für unterschiedliche Themen verschiedene Ansprechpartner. Im Internet können sie sich darüber hinaus auch an Unbekannte richten. Der Anreiz besteht darin, die passenden Menschen zusammenzubringen. Im Folgenden werden Funktionsweisen eines einzelnen Gehirns auf ein Globales Gehirn übertragen, in dem der einzelne Mensch wie eine Zelle in einem intelligenten Gesamtsystem fungiert. Ziel ist es, durch strukturierte Verknüpfungen eine koordinierte Kommunikation zwischen passenden Partnern zu ermöglichen.

3.1 Das Gehirn

Lebewesen nehmen mit ihren Sinnen Informationen auf, verarbeiten diese und reagieren. Solche Reiz-Reaktions-Ketten sind essentiell für Selbst- und Arterhalt. Sinneszellen übersetzen einen empfangenen Reiz in eine Erregung, die von Neuronen weitergeleitet und verarbeitet wird. Muskel- oder Drüsenzellen sind die reagierenden Organe.

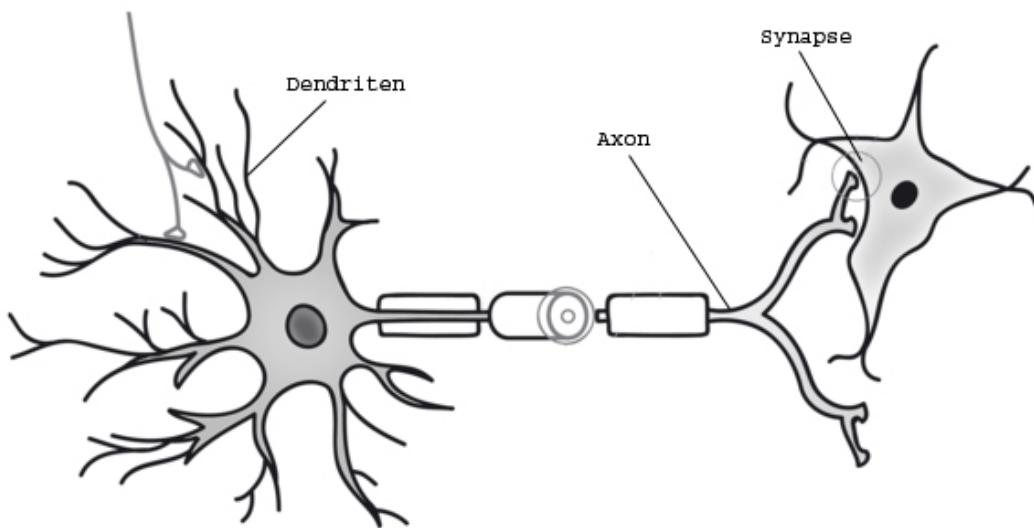


Abbildung 7: Verbindung zweier Neuronen⁵³

Neuronen sind für die Weiterleitung und Verarbeitung von Reizen zuständig. Sie empfangen Signale von Sinneszellen oder anderen Neuronen und leiten diese über das Axon weiter. Durch die Ausschüttung von Botenstoffen werden Signale an nachgeschaltete Zellen weitergeleitet.

Neuronen sind elektrisch erregbar. Sind sie nicht erregt, befinden sie sich im Ruhepotential, das ist ein Ladungsunterschied zwischen Zellinnerem und -äußerem. Wird ein ruhendes Neuron gereizt, verringert sich dieser Ladungsunterschied. Falls dabei ein Schwellwert überschritten wird, folgt darauf eine kurze Umpolung, also ein Aktionspotential. Dieses wird über das Axon weitergeleitet, um nachgeschaltete Zellen zu informieren.

53 Bildquelle: <http://de.wikipedia.org/wiki/Bild:Nervenzelle.png>

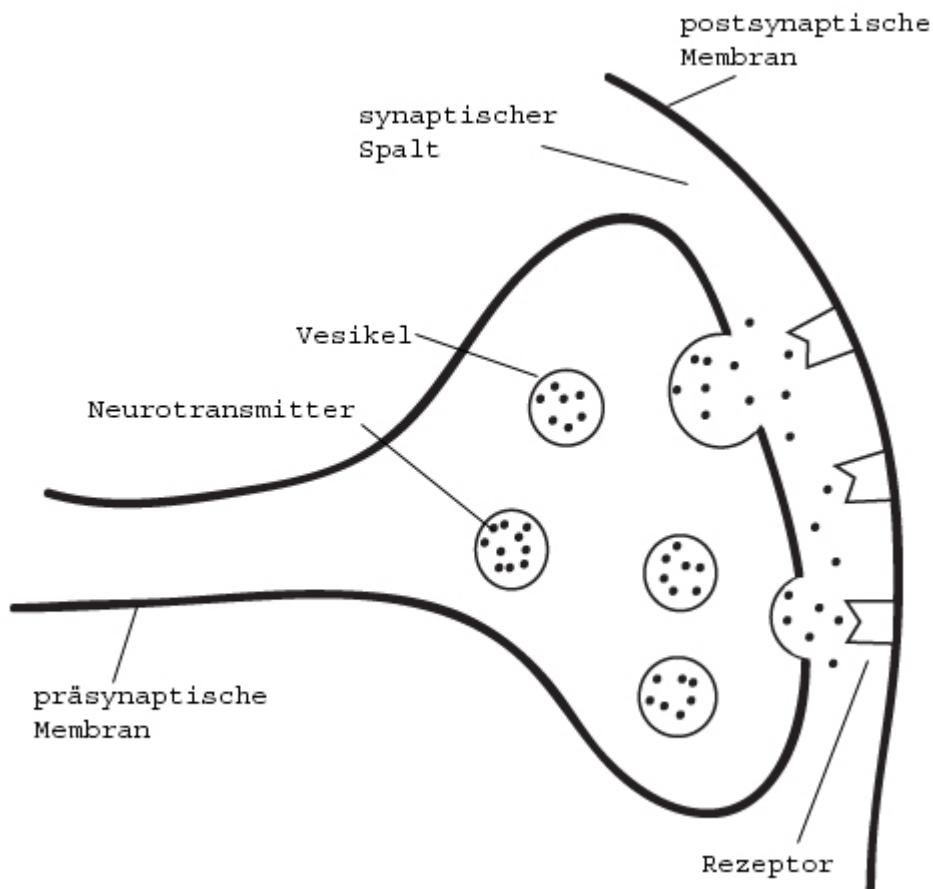


Abbildung 8: Schematische Darstellung einer Synapse

Die Schalt- und Verbindungsstellen zwischen Neuronen heißen Synapsen und dienen der Erregungsübertragung und Informationsverarbeitung. Zwischen dem axonalen Endknöpfchen und der Membran der benachbarten Zelle befindet sich ein schmaler Abstand, der synaptische Spalt, den Aktionspotentiale nicht direkt überspringen können.

Die Erregungsübertragung erfolgt über chemische Botenstoffe, die Neurotransmitter genannt werden und in den synaptischen Vesikeln gespeichert sind. Ein am Endknöpfchen ankommendes Aktionspotential veranlasst das Verschmelzen der Vesikel mit der Zellmembran, so dass die Transmitter in den synaptischen Spalt diffundieren und dort die Rezeptoren der postsynaptischen Membran erreichen, an die sie sich binden. Erst eine gewisse Frequenz von Aktionspotentialen in der Präsynapse und damit die Ausschüttung von genügend Transmitterstoffen löst in der Postsynapse ein Aktionspotential aus. Es gibt auch hemmende Synapsen, die die Auslösung eines Aktionspotentials erschweren. Im Gehirn senden oft mehrere Tausend Neurone ihre Signale zu einem nachgeschalteten Neuron, das entsprechend der Gesamtsumme dieser Signale reagiert.

Das Gehirn ist ein riesiges Netzwerk aus Neuronen. Beispiele für einfache Schaltungen sind Reihen, Ringe, Divergenz und Konvergenz. Die Leistungsfähigkeit von Synapsen und Schaltungen wird durch häufige Nutzung gesteigert, da jedes Feuern (jede

Erregungsübertragung) die entsprechende Synapse festigt. Solche langfristigen Veränderungen am Neuronennetz scheinen die Basis des Langzeitgedächtnisses zu sein.

In der Großhirnrinde sind sensorische Felder für die Wahrnehmung, motorische Felder zur Steuerung der Bewegung und wenig erforschte assoziative Felder zu erkennen.

Die Funktionsprinzipien des menschlichen Gehirns dienen auf dem Forschungsgebiet der Künstlichen Intelligenz als Vorlage, um intelligentes Verhalten zu erzielen. Durch das Zusammenschalten mehrerer neuronaler Einheiten entstehen neuronale Netze. Diese dienen der Mustererkennung und sollen durch wenige ausgewählte Eingabemuster lernen, eine beliebige Eingabe in das gewünschte Ausgabemuster zu überführen.

Neuronale Einheiten werden mit gewichteten Eingängen modelliert, die sich zusammen zum Nettoinput summieren, der über eine Schwellwertfunktion zu einem Ausgabewert umgewandelt wird. Neuronale Netze unterscheiden sich hinsichtlich ihrer Neuronenfunktion, Topologie (Struktur der Verbindungen) und der Veränderung der Gewichte zwischen den Neuronen, die durch Lernregeln gesteuert wird. Die einfachste Lernregel ist die Hebb'sche Regel, die die Tatsache beschreibt, dass die Änderung der Verbindungsstärke zwischen den sendenden und einem empfangenden Neuron von dem Input der sendenden Neuronen und der Aktivierung des empfangenden Neurons abhängt.

3.2 Die Vernetzung mehrerer Gehirne

We are forming cells within a global brain and we are excited that we might start to think collectively. What becomes of us still hangs crucially on how we think individually.⁵⁴

Im Gegensatz zu neuronalen Netzen sind im Global Brain Menschen die einzelnen Zellen, die durch den Austausch von Nachrichten individuell miteinander verbunden sind. Das Global Brain soll es Menschen ermöglichen, Nachrichten an Unbekannte zu versenden, die sich nur anhand ihrer gewünschten Eigenschaften und den Themen, für die sie sich interessieren, charakterisieren lassen. Damit entsteht ein virtuelles Modell für die Zusammenarbeit von Menschen im Internet.

Analoge Reize, die wir im wirklichen Leben erfahren, müssen in eine digitale Repräsentation transformiert werden. Dabei lassen sich alle multimedialen Formate wie Bild, Text, Ton und Video verwenden, die auch heute schon "wirkliche Gefühle" vermitteln und Reaktionen auslösen können.

Es gibt verschiedene Motivationen, eine Nachricht zu verfassen, wie zum Beispiel eine Frage zu stellen oder ein Inserat aufzugeben. Nachrichten sind mit Zeit- und

54 Realising the Full Potential of the Web, Tim Berners-Lee: <http://www.w3.org/1998/02/Potential.html>

Kostenkoeffizienten verbunden. Die Gültigkeitsdauer einer Nachricht kann schnell ablaufen oder lange anhalten. Der Wert einer Nachricht kann positiv oder negativ sein. Entweder, der Urheber bezahlt dafür, dass jemand seine Nachricht wahrnimmt und beantwortet, wie bei der Werbung, oder der Empfänger bezahlt dafür, die Nachricht lesen zu dürfen, wie es zum Beispiel bei redaktionellen Inhalten der Fall ist. Der Empfänger selektiert eingehende Nachrichten, interpretiert sie und kann sie verwerfen oder beantworten.

Um Nachrichten beschreiben, adressieren und vermitteln zu können, ist eine semantische Abstraktion nötig. Die Beschreibung einer Nachricht umfasst mehrere Merkmale. Ob eine Nachricht vermittelt wird, hängt von der Gesamtheit der ausgewählten Merkmale sowohl auf der sendenden als auch auf der empfangsbereiten Seite ab.

Nachrichten sollen abhängig von ihrem Inhalt entweder sehr gezielt an passende Stellen übermittelt oder an einen weiten Empfängerkreis verteilt werden. Auf der anderen Seite muss es jedem Empfänger möglich sein, für sich passende Nachrichten gezielt herauszufiltern oder selbst aus einer großen Menge von Nachrichten zu selektieren. Dies entspricht den Mechanismen von Synapsen und Dendriten.

So wie die Vermittlung im Gehirn über die Synapsen durch chemische Neurotransmitter und Rezeptoren erfolgt, braucht die globale Vermittlung an jedem einzelnen Merkmal einen Steuermechanismus.

Einen Vorschlag, diesen zu realisieren, bietet OLI-it. Auf der sendenden und empfangsbereiten Seite werden für jedes Merkmal Regeln verwendet. Dadurch ist es beispielsweise möglich, festzulegen, dass ein bestimmtes Merkmal sehr wichtig ist und dass die andere Seite es deshalb auch ausgewählt haben muss.

Wenn wir uns vorstellen, dass Menschen durch das Global Brain miteinander verbunden sind, bedeutet dies einen evolutionären Schritt für die sozialen Verflechtungen der Gesellschaft und eine Bereicherung für jeden Einzelnen.

4 Beschreibung von OLI-it

OLI-it ist eine Idee von Frederic Luchting. Es handelt sich hierbei um eine Kommunikationsinfrastruktur, die es ermöglicht, Nachrichten zwischen Unbekannten zu vermitteln und die Vision des Global Brain zu realisieren. Im Folgenden werden die einzelnen Komponenten von OLI-it beschrieben.

4.1 Kreislauf

Der Kreislauf geht von der Erzeugung einer Nachricht bis zu ihrer Beantwortung. Ein einfacher Ablauf der Kommunikation im Internet ist folgende Kette:

Sender -> Nachricht -> Empfänger -> Antwort -> Bewertung -> Belohnung

4.1.1 Elemente

Bei OLI-it werden für Kernelemente des Kreislaufs eigene Namen verwendet.

- Stamm
Ein Stamm ist ein Anwender, also ein Mensch oder ein Agent, der Inhalte produziert. Dieser kann Urheber oder Empfänger einer Nachricht sein.
- PostIt
Ein PostIt ist eine Nachricht. Hierbei kann es sich um alles Erdenkliche handeln, wie beispielsweise um eine Frage, eine Kleinanzeige oder die Suche nach einem passenden Partner.
- TopLab
Ein TopLab ist die Antwort auf eine Nachricht. Es kann bewertet, kommentiert und belohnt werden.

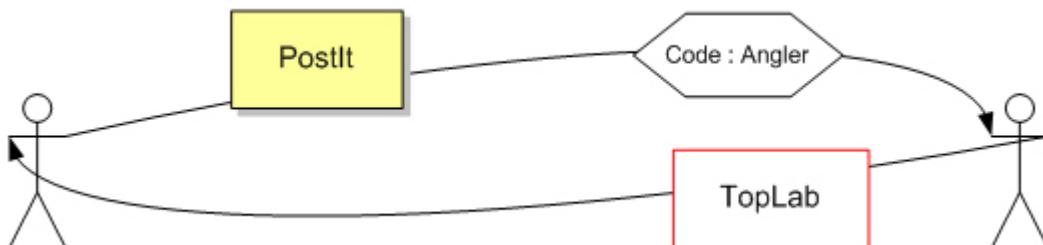


Abbildung 9: Der Kreislauf von OLI-it

Ein Stamm versendet sein PostIt an geeignete Empfänger, die möglicherweise mit einem TopLab antworten. In entgegengesetzter Richtung fließen Punkte als symbolisches Geld. Der Urheber versieht sein PostIt mit einem positiven oder negativen Betrag, den er nach Ablauf der von ihm gesetzten Frist unter den Antworten aufteilt. Damit wird OLI-it den Zeit- und Kostenkoeffizienten von Nachrichten gerecht.

Nun geht es um die Vermittlung eines PostIt an die passenden Empfänger. Dazu markiert ein Stamm in der Rolle des Urhebers sein PostIt in einem semantischen Wortraum mit einem so genannten Code. Ein Stamm in der Rolle eines potentiellen Empfängers markiert einen so genannten Angler, um passende PostIts zu finden. Code und Angler werden abgeglichen; falls sie zueinander passen, wird das PostIt an den Empfänger vermittelt.

4.2 Wortraum

Die Beschreibung des Inhalts einer Nachricht, ihres Urhebers und der gewünschten Empfänger erfolgt in einem erweiterbaren Wortraum. Hier werden relevante Merkmale wie in einem Koordinatensystem markiert.

Der Wortraum ist eine teilweise hierarchische Struktur. Er besteht aus Netzen und Bäumen, die Hilfsstrukturen für Knoten und Zweige sind:

Netz mit Knoten



Ein Netz gruppert unter einem Oberbegriff mehrere Bereiche, die als Knoten bezeichnet werden.

Baum mit Zweigen



Ein Baum ist ein Verzweigungspunkt in einer strengen Hierarchie, also eine Sammlung disjunkter Elemente, die als Zweige bezeichnet werden.

Knoten und Zweige können wieder zu Netzen und Bäumen weiterführen. Die Beziehungen zwischen Netz und Knoten bzw. Baum und Zweig sind Kompositionen, das bedeutet, dass ein Knoten Teil genau eines Netzes und ein Zweig Teil genau eines Baumes ist.

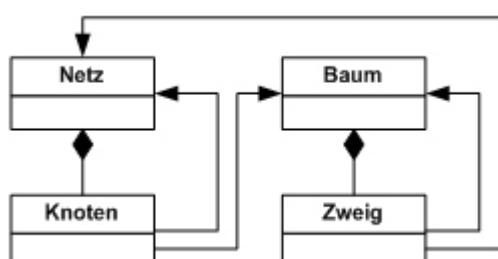


Abbildung 10: Elemente des Wortraums

Eine mit Hilfe eines Baumes beschriebene Hierarchie kann im Gegensatz zu einem Netz mehrfach verwendet werden.

Dadurch, dass die einzelnen Merkmale, also die Knoten und die Zweige, in einer semantischen Wissensstruktur eingeordnet sind, erhalten sie einen eindeutigen Kontext.

4.3 Vermittlungslogik

Die Vermittlungslogik von OLI-it besteht aus den Markierungen von Code und Angler, die weitere Elemente des Kreislaufs sind, und deren Abgleich.

4.3.1 Markierung

Um PostIts zu beschreiben und eine gezielte Vermittlung zu ermöglichen, werden passende Merkmale im Wortraum mit einem Code markiert. Analog dazu wird eine Suchanfrage mit einem Angler markiert. Jede dabei ausgewählte Koordinate wird mit zwei Werten versehen, die den Abgleich gegen die gegenüberliegende Seite steuern.

- **Code**

Ein Code ist eine Menge von Markierungen eindeutiger Merkmale des Wortraums mit zwei Werten, die den Abgleich gegen Angler steuern. Jede Koordinate im Wortraum wird durch ein Quadrupel, bestehend aus Netz, Knoten, Baum und Zweig, identifiziert. Da auf dem Weg zu spezielleren, also tiefer liegenden Merkmalen im Wortraum weniger spezielle Merkmale auf höheren Ebenen markiert werden, ist ein Code ein Teilbaum des Wortraums, in dem jeder Knoten zusätzlich mit zwei Werten versehen ist.

- **Angler**

Ein Angler ist eine Menge von Markierungen eindeutiger Merkmale des Wortraums mit zwei Werten, die den Abgleich gegen Codes, also die Markierungen von PostIts, steuern.

Codes und Angler werden von der Wurzel des Wortraums ausgehend markiert. Auf der obersten Ebene befinden sich drei Knoten für Urheber, Nachricht und Empfänger. Von diesen aus kann immer spezieller und detaillierter weitere Markierungen vornehmen.

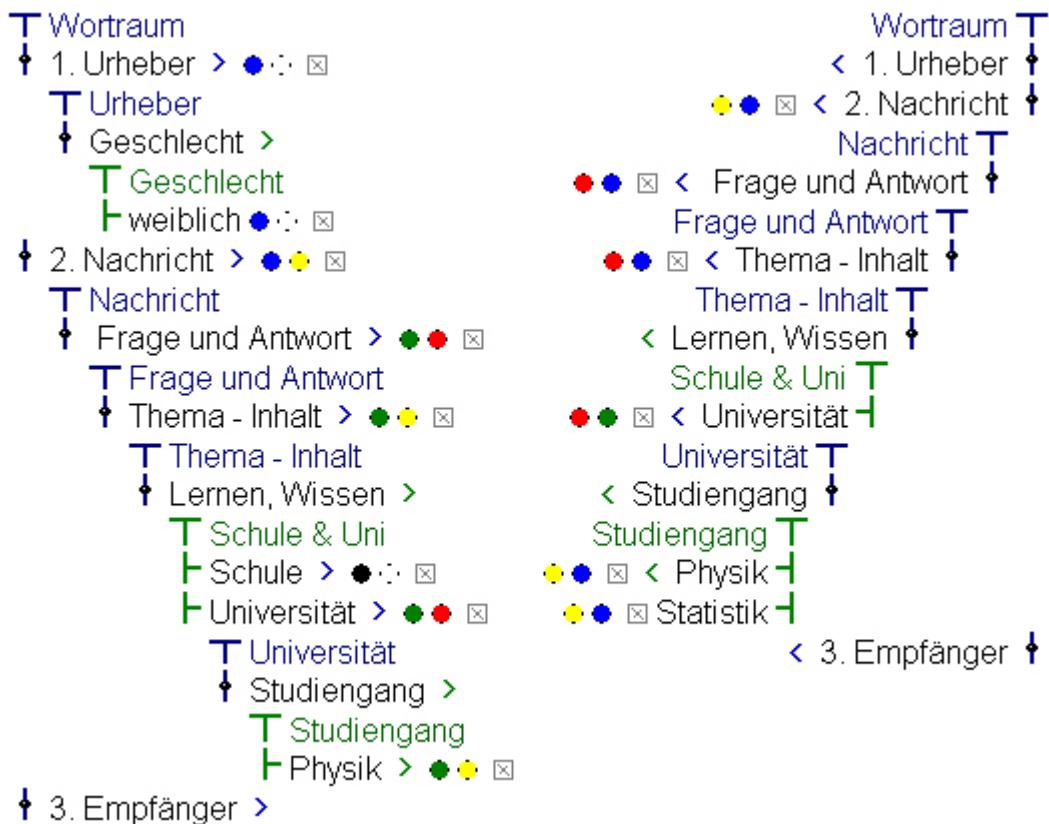


Abbildung 11: Ein Code mit passendem Angler

Die Abbildung zeigt den Code einer Nachricht mit einem passenden Angler. Die Urheberin gibt an, dass es sich bei dem Inhalt ihrer Nachricht um eine Frage zum Thema Physik handelt. Der Empfänger der Nachricht interessiert sich für Fragen zum Thema Physik oder Statistik und kann die Frage eventuell beantworten.

4.3.2 Werte

Jede ausgewählte Koordinate eines Codes oder Anglers wird mit zwei Werten markiert. In einem Code heißen die beiden Werte OLIs und get, in einem Angler heißen sie ILOs und fit. Im Folgenden wird auf die beiden Werte der Codes eingegangen, die sich auf die ILOs der Angler beziehen. Die beiden Werte der Angler beziehen sich auf die OLIs der Codes und funktionieren genau gleich.

Mit den OLIs wird festgelegt, wie abgeglichen wird. Es stehen drei unterschiedliche Werte zur Verfügung:

- 3:

Der Wert 3 repräsentiert ein sehr strenges Kriterium. Beim Markieren darf der Wert 3 innerhalb einer Struktur, also innerhalb eines Netzes oder Baumes, nur einmal gewählt werden (Einfachauswahl). Wird dieser Wert gewählt, muss die Koordinate auf der gegenüberliegenden Seite die Bedingung des get erfüllen.

- 2:

Der Wert 2 repräsentiert ein weniger strenges Kriterium und darf deshalb innerhalb einer Struktur mehrfach verwendet werden (Mehrfachauswahl). Wird dieser Wert gewählt, muss für mindestens eine Koordinate pro Struktur gelten, dass auf der gegenüberliegenden Seite die Bedingung des get erfüllt wird.

- 1:

Der Wert 1 bedeutet, dass die Koordinate auf der Anglerseite nicht markiert sein darf und verhindert andernfalls eine Vermittlung.

Mit get wird bestimmt, wie die Markierung auf der gegenüberliegenden Seite aussehen soll. Es stehen die Werte 3, 2 und 0 zur Verfügung. Ein Angler wird bei der Vermittlung nur dann berücksichtigt, wenn der Wert des ILOs größer oder gleich dem Wert des get ist. Der Wert 0 steht für eine freiwillige Markierung, deshalb muss die Koordinate auf der gegenüberliegenden Seite nicht markiert sein.

Damit ergeben sich folgende sinnvolle Kombinationen:

OLIs	get	Beispiel	Erklärung
3	3	Frage und Antwort  	Die strengste Markierung wird für eigenständige Teilbereiche verwendet, um sie eindeutig von anderen Bereichen abzugrenzen.
3	2	Physik  	Um auch Empfänger zuzulassen, die eine Mehrfachauswahl getroffen haben, hat get den Wert 2.
2	2	Nachricht  	Mit dieser Markierung kann man selbst eine Mehrfachauswahl vornehmen, von der mindestens ein Merkmal erfüllt sein muss.
2	0	weiblich  	Wenn jemand freiwillig ein Merkmal angibt, kann er dadurch gefunden werden, stellt aber keine Bedingungen an den Suchenden.
1		Schule  	Empfänger, die "Schule" markiert haben, werden ausgeschlossen.

4.3.3 Abgleich

Alle Koordinaten, die markiert wurden, ergeben zusammen den Code eines PostIt bzw. einen Angler. Der Abgleich der Codes mit den Anglern muss gewährleisten, dass die in der Markierung angegebenen Regeln für beide Seiten erfüllt sind.

Der ALGORITHMUS des Abgleichs prüft für jedes OLI mit dem Wert 3, ob auf der Anglerseite die entsprechende Koordinate mit einem Wert markiert wurde, der größer oder gleich dem get ist. Ist dies für eine Markierung nicht der Fall, wird der Algorithmus abgebrochen, da Code und Angler nicht zusammenpassen.

Andernfalls werden alle OLIs mit dem Wert 2 überprüft. Hierbei genügt es, wenn für nur ein Element in seiner übergeordneten Struktur auf der Anglerseite eine Markierung gefunden wird, die die Bedingung des get erfüllt.

Zuletzt wird für jedes OLI mit dem Wert 1 sichergestellt, dass die entsprechende Koordinate auf der Anglerseite nicht markiert wurde.

Analog dazu wird der Angler gegen den Code abgeglichen. Sind auf beiden Seiten alle Bedingungen erfüllt, wird das PostIt vermittelt.

4.4 Realisierung

OLI-it⁵⁵ wurde als datenbankgestützte Webanwendung entwickelt.

Für alle vorgestellten Komponenten existieren in der Datenbank Tabellen, die konkrete Instanzen enthalten. Diese werden über GUIDs als Primärschlüssel eindeutig bezeichnet.

Der Datenzugriff, die Datenmanipulation und die Anwendungslogik sind von einer Mittelschicht gekapselt, die alle Funktionen auf einer abstrakten Ebene in Form eines API zur Verfügung stellt. Ein API bietet eine universelle Möglichkeit, beliebige Anwendungen, wie beispielsweise Desktopanwendungen, mobile Applikationen und Weboberflächen darauf aufzubauen.

Es existiert eine Referenzimplementierung als Webanwendung. Diese bietet die Möglichkeit, konkrete Anwendungen wie Jobbörsen, Frage-und-Antwort-Systeme sowie Partnervermittlungsportale über den Nachrichtenaustausch zu realisieren. Darüber hinaus existieren für eine Webanwendung übliche Funktionen wie Personalisierung, automatische E-Mail-Benachrichtigung, Bewertung und Belohnung.

55 OLI-it: <http://www.oli-it.com>

5 Beurteilung von OLI-it

Anschließend werden die Vor- und Nachteile sowie die Einsatzmöglichkeiten von OLI-it genauer beleuchtet.

Die vorgestellten Ansätze bieten einer großer Anzahl potentieller Empfänger eine Masse von Nachrichten, aus denen diese sich selbst die für sie interessanten selektieren können.

Selektionsdruck

Als Suchender nach konkreten Antworten findet man mit Suchmaschinen oder auf Wikis mit Glück eine gut erklärte Zusammenfassung. In Newsgroups muss man eventuell ganze Threads, die für Außenstehende allein wegen ihrer Länge manchmal nicht leicht nachzuvollziehen sind, durchforsten, bis man das gewünschte Ergebnis findet.

5.1 Beurteilung von OLI-it anhand der vorgestellten Kriterien

Asynchron

many-to-many

offen

Benutzername

Webanwendung

neu

RDF

offen

ja

ja

5.2 Nachteile

Einer der auffälligsten Nachteile von OLI-it ist die für neue Benutzer schwierige Terminologie. Deshalb ist der erste Verbesserungsvorschlag, diese vor dem Benutzer zu verbergen und die einzelnen Elemente mit eindeutigen, intuitiv erfassbaren Namen zu

versehen. Beispielsweise könnte ein Stamm als User, ein PostIt als Nachricht, ein TopLab als Antwort, ein Code als Markierung und ein Angler als Filterprofil präsentiert werden.

Auch der Wortraum ist genauer zu beleuchten. Hier geht es zunächst um die Strukturen Netz und Baum, deren Namen wiederum verwirrend sind. Sie unterscheiden sich vor allem dadurch, dass Bäume im Gegensatz zu Netzen mehrfach verwendet werden können. Auf der obersten Ebene des Wortraums befinden sich die beiden Netze Urheber und Empfänger, die beide dieselben Knoten enthalten.

Insgesamt stellt sich die Frage, ob eine Ontologie oder ein semantisches Netz besser geeignet wären, um Metabeschreibung für Nachrichten zu erstellen. Der Wortraum ist eine Taxonomie und damit eine einfache Ontologie. In einer Ontologie oder einem semantischen Netz könnte man nicht ohne Weiteres die Markierungen übernehmen, da sich diese auf die Strukturen beziehen.

Auch die beiden Werte für die Markierung sind kompliziert. Da es nur 5 verschiedene Kombinationsmöglichkeiten gibt, könnte man diese beispielsweise auch durch die Ziffern 1 - 5 bezeichnen.

Insgesamt ist für eine so komplizierte Anwendung viel Unterstützung für den Benutzer erforderlich. So wären einige Tools wünschenswert, die dem Benutzer Arbeit abnehmen, wie zum Beispiel ein automatisches oder teilweise automatisiertes Markieren, hoch markieren, runter löschen, natürliche Spracherkennung, Inhalt indexieren...

Ein einfacher Schritt wäre es, das Benutzerprofil beim Anmelden direkt abzufragen, um es dann immer wieder verwenden zu können.

5.3 Vorteile

Matchmaking zwischen Menschen: Verbindungen erarbeiten, die der Benutzer noch nicht kennt, beispielsweise das Finden von anderen Personen in einer Organisation, die sich für ein bestimmtes Thema interessieren. Jemanden finden, den man nach bestimmten Informationen fragen kann. Direkten Kommunikationskanal zu einer anderen Person aufbauen.

Menschen erreichen anhand ihrer Lokalität und Ihres Berufs, große Datenmengen verwalten mächtig, kann NG und Foren ersetzen

Man kann Menschen anhand ihrer Eigenschaften erreichen (beispielsweise jemanden, der in Hagen wohnt) und kann durch die fünf verschiedenen Markierungsmöglichkeiten feiner abgleichen als durch übliche und - oder - nicht Verknüpfungen.

Positiv ist zudem, dass jeder seine Kriterien und die Regeln, nach denen diese abgeglichen werden, selbst bestimmen kann.

Durch den gezielten Austausch von Mensch zu Mensch entsteht ein spannendes und lebendiges System.

Durch die Kombinationsmöglichkeiten der Merkmale ergeben sich einzigartige Beschreibungen, die sehr individuell abgefragt werden können. Die erzielten Treffer passen inhaltlich besser als bei einem flachen Wortabgleich überein. Die Interessen von Urheber und Empfänger werden gleichermaßen berücksichtigt: durch ihre individuelle Markierung erreicht eine Nachricht nur gewollte Empfänger, die ihrerseits auch nur gewünschte Nachrichten empfangen.

Ein weiterer Vorteil ist die Möglichkeit, eine sehr detaillierte und freiwillige Markierung, wie zum Beispiel die genaue Angabe einer Lokalität, die tief in die Hierarchie hinabreicht, auf beliebig höheren Ebenen abzufragen. Ist etwa Hagen markiert, kann auch jemand, der nur in Nordrhein-Westfalen sucht, die Nachricht bekommen, da auf dem Weg nach Hagen auch Nordrhein-Westfalen markiert wurde.

Alter Vergleich:

Usenet:

Auch OLI-it nutzt einen hierarchischen Wortraum, um Nachrichten zu markieren. Nachrichten liegen jedoch nicht in ihrer Kategorie, sondern auf einem "Haufen" und sind durch Markierungen gekennzeichnet, die sich an eine eindeutige Kombination mehrerer Kategorien richten können. Das hat den Vorteil, dass die Kategorien nicht fest vorgegeben sind und Nachrichten sehr individuell eingestellt und gesucht werden können.

Lycos:

Dies scheint eine große Ähnlichkeit mit OLI-it zu haben. Unterschiede liegen vor allem in der Vermittlung: bei Lycos iQ bekommt man alle Fragen vorgeschlagen, zu denen mindestens ein Tag mit einem eigenen übereinstimmt. Es gibt keine Möglichkeit, individuellere Regeln für die Vermittlung festzulegen. OLI-it bietet darüber hinaus die Möglichkeit, Nachrichten an Empfänger mit gewünschten persönlichen Eigenschaften zu adressieren, die ihrerseits nach Nachrichten von Urhebern mit gewünschten Eigenschaften suchen können.

Online Marktplätze:

OLI-it hat diesen Systemen gegenüber den Vorteil, dass Kriterien frei gewählt und individuell miteinander abgeglichen werden können.

Services für Inhalte: Vermitteln, Suchen, Darstellen wie Google oder Technokrati

6 RDF Implementierung von OLI-it

Das Semantic Web und OLI-it ergänzen sich symbiotisch. Das Semantic Web gewinnt durch OLI-it die Möglichkeit, Menschen über Nachrichten miteinander zu verbinden. OLI-it gewinnt mit dem Semantic Web neue Technologien, die Dezentralisierung, Verteilbarkeit und Unabhängigkeit ermöglichen.

Im Folgenden wird beschrieben, wie diese beiden Ideen zusammengeführt werden, um Nachrichten im Semantic Web zu vermitteln. Natürlich lassen sich auch die oben beschriebenen alternativen Anwendungen auf die Technologien des Semantic Web übertragen. Jedoch bietet keine wie OLI-it die Möglichkeit einer globalen und individuellen Vermittlung, die auf Regeln basiert.

Damit andere Anwendungen wie beispielsweise intelligente Softwareagenten OLI-it im Semantic Web nutzen können, wird das Anwendungsmodell mit RDF Schema formalisiert und als Ontologie zur Verfügung gestellt. Konkrete Daten werden im RDF Format unter Verwendung dieses RDF Schemas über Webschnittstellen für Menschen und Web Services für Clients importiert und exportiert. Auch der Abgleich zwischen Codes und Anglern wird über Web Services zur Verfügung gestellt, damit nicht jede Anwendung die Abgleichslogik von OLI-it implementieren muss. Zuletzt wird ein Client als Referenzimplementierung vorgestellt, der nur über Web Services mit OLI-it kommuniziert und dazu dient, neue Inhalte zu erstellen und abzugleichen.

6.1 RDF Schema

Um die Daten von OLI-it interpretieren und auf weitere Informationen schließen zu können, wird ein gemeinsames RDF Schema benötigt, das die Struktur der Elemente des Kreislaufs, des Wortraums und der Vermittlungslogik beschreibt.⁵⁶

Dazu wurden Tabellen der Datenbank als Klassen und die Beziehungen zwischen diesen als Eigenschaften modelliert.

56 Das vollständige RDF Schema liegt unter <http://nulllogicone.net/schema.rdfs>

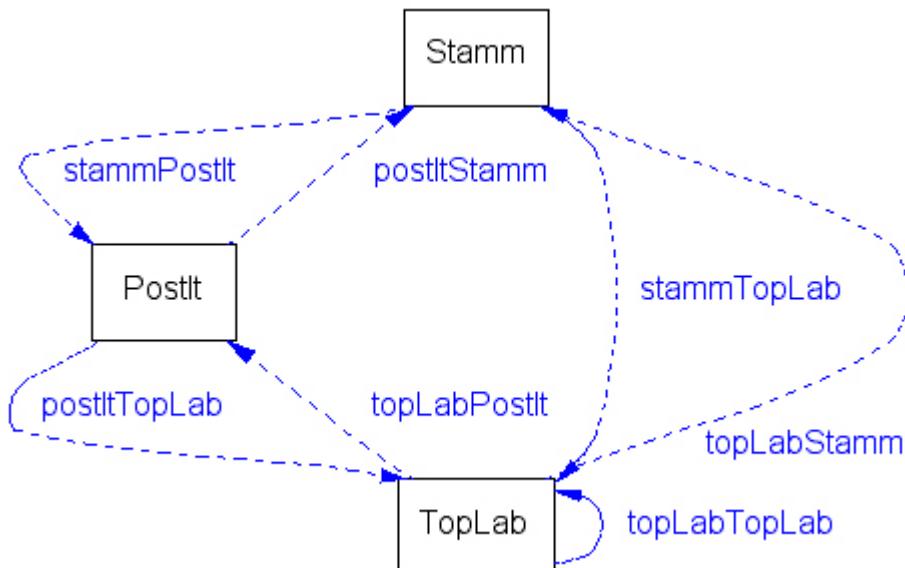


Abbildung 12: Beziehungen zwischen Klassen⁵⁷

Die Graphik zeigt Beziehungen zwischen drei ausgewählten Klassen des Kreislaufs. Auf ein PostIt kann es mehrere TopLabs geben. Ein Stamm kann mehrere PostIts schreiben und ein PostIt kann zu mehreren Stämmen gehören. Außerdem kann man TopLabs auf TopLabs abgeben. Diese Beziehungen wurden in dem RDF Schema durch die Angaben von Domain und Range definiert.

Folgender Ausschnitt aus dem Schema zeigt nur die Klasse Stamm mit ihren Eigenschaften:

⁵⁷ Das RDFS Dokument liegt unter <http://www.stud.fernuni-hagen.de/q6261035/Bachelorarbeit/SPT.rdfs>, die Graphik wurde mit dem FRODO RDFSviz Tool erstellt: <http://www.dfki.uni-kl.de/frodo/RDFSviz>

```

<rdf:RDF xml:lang="de"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:base="http://nulllogicone.net/schema.rdfs#">

  <!-- Klasse Stamm -->
  <rdfs:Class rdf:ID="Stamm"></rdfs:Class>

  <!-- Stamm Eigenschaften -->
  <rdf:Property rdf:ID="stammPostIt">
    <rdfs:domain rdf:resource="#Stamm"/>
    <rdfs:range rdf:resource="#PostIt"/>
  </rdf:Property>

  <rdf:Property rdf:ID="stammTopLab">
    <rdfs:domain rdf:resource="#Stamm"/>
    <rdfs:range rdf:resource="#TopLab"/>
  </rdf:Property>

</rdf:RDF>

```

Für die Abgleichslogik ist es wichtig, dass die Werte der Markierungen in Codes und Anglern nur definierte Zustände annehmen können. Deshalb wurden sie im RDF Schema als Subklassen ihrer Typen modelliert.

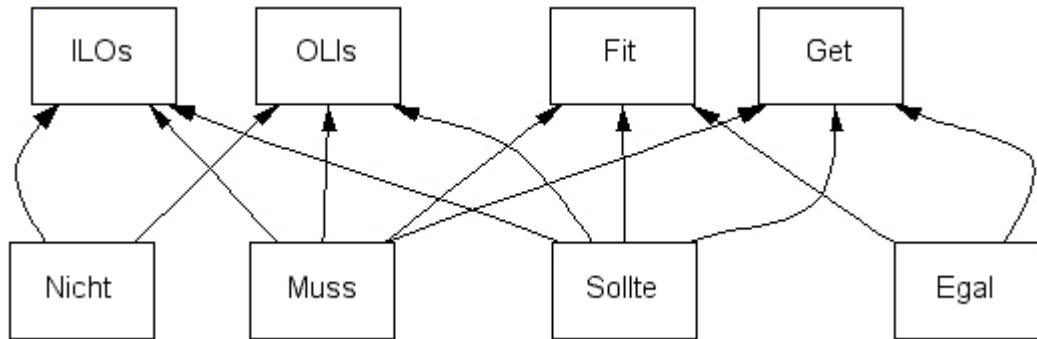


Abbildung 13: Werte für eine Markierung⁵⁸

Die Namen der Subklassen stehen für die verschiedenen Werte der Markierungen. Da Klassen nicht mit Ziffern benannt werden dürfen, steht "Muss" für den Wert 3, "Sollte" für den Wert 2, "Nicht" für den Wert 1 und "Egal" für den Wert 0.

6.2 RDF

Neuigkeiten von OLI-it werden als RSS Feeds zur Verfügung gestellt, um Interessierten die Möglichkeit zu bieten, informiert zu bleiben, ohne eine Webseite immer wieder direkt

⁵⁸ Die Graphik wurde mit dem FRODO RDFS Viz Tool erstellt: <http://www.dfki.uni-kl.de/frodo/RDFSViz>

aufrufen zu müssen oder durch E-Mail-Benachrichtigungen belästigt zu werden. Zum Beispiel liefert das Feed <http://xml.oli-it.com/RSS/StammPostIt.aspx?sguid=479bbe9a-6fbf-4acb-b604-d9e34fa955f0> die neuesten Einträge des MetaWeb Project.⁵⁹

Werden die Daten als RDF Statements mit zugehörigem RDF Schema formuliert, eröffnet sich die Möglichkeit, mit einer Abfragesprache wie SPARQL Anfragen an Aussagen zu stellen. Um diesen und weitere Vorteile von RDF zu nutzen, wurde die Ausgabe sämtlicher Inhalte der Datenbank, also Stamm, Angler, PostIt, Code und TopLab, als RDF Dokumente ermöglicht.

Zu jedem einzelnen Datensatz gibt es ein RDF Dokument, das unter einem URL nach folgendem Schema aufgerufen werden kann:

`http://nulllogicone.net/Tabellenname/GUID.rdf`

Die Relationen zwischen den Datensätzen werden in Statements ausgedrückt, die zu weiteren Dokumenten führen. Damit lässt sich der Datenbestand traversieren.

Jedes dieser Dokumente hat den Datensatz, den es beschreibt, als Subjekt. Der URI des Subjekts hat folgende Form:

`http://nulllogicone.net/Tabellenname/?GUID`

Mit dieser URI als Adresse in einem Browser kann man den zugehörigen Datensatz sehen und zu weiterführenden Datensätzen navigieren.

Folgendes Beispiel zeigt ein PostIt mit einem seiner Urheber:

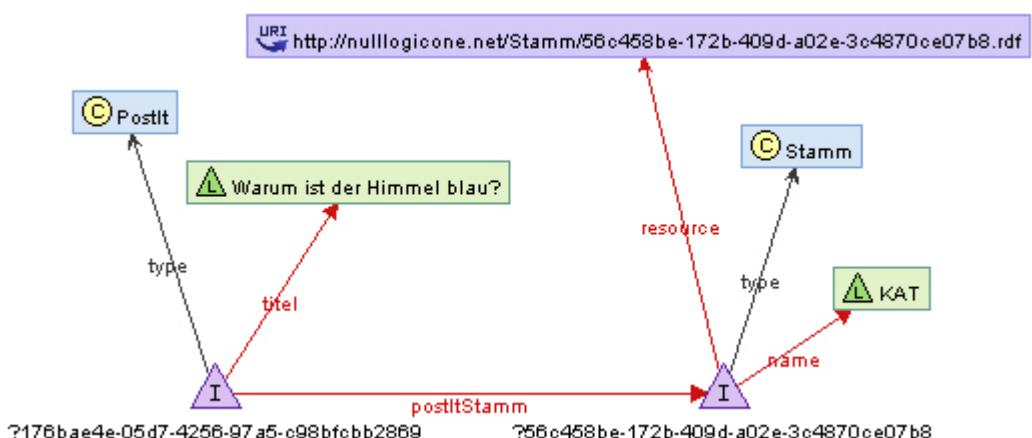


Abbildung 14: Ein PostIt mit seinem Stamm⁶⁰

59 The MetaWeb Project, eine spezialisierte Anwendung von OLI-it, um Homepages anzumelden, zu beschreiben und zu finden: <http://www.metaweb-project.net>

Dieser Graph wird in RDF/XML folgenderweise serialisiert:

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:nlo="http://nulllogicone.net/schema.rdfs#"
  xml:base="http://nulllogicone.net/">

  <nlo:PostIt
    rdf:about="PostIt/?176bae4e-05d7-4256-97a5-c98bfcb2869">
    <nlo:titel>Warum ist der Himmel blau?</nlo:titel>

    <nlo:postItStamm>
      <nlo:Stamm
        rdf:about="Stamm/?56c458be-172b-409d-a02e-3c4870ce07b8"
        nlo:name="KAT">
        <nlo:resource rdf:resource="http://nulllogicone.net/Stamm/
          56c458be-172b-409d-a02e-3c4870ce07b8.rdf"/>
      </nlo:Stamm>
    </nlo:postItStamm>

  </nlo:PostIt>
</rdf:RDF>61
```

Das PostIt wird mit dem URI <http://nulllogicone.net/PostIt/?176bae4e-05d7-4256-97a5-c98bfcb2869> bezeichnet.

Die vollständige RDF Datei liegt unter: <http://nulllogicone.net/PostIt/176bae4e-05d7-4256-97a5-c98bfcb2869.rdf>

Um die RDF Dokumente mit der Datenbank von OLI-it austauschen zu können, wurde eine Schnittstelle für den Im- und Export⁶² von Daten programmiert. Zu den GUIDs von Datensätzen werden RDF Dokumente ausgegeben. Auch der komplette Wortraum lässt sich als RDF Dokument exportieren. RDF Dokumente für Code und Angler können geparsed, validiert und in der Datenbank gespeichert werden. Für valide RDF Dokumente werden eine Tabellenansicht, eine Wortraumansicht und die passenden Angler bzw. Nachrichten angezeigt. Durch den Im- und Export wird eine dezentrale Haltung und Bearbeitung der Daten ermöglicht.

60 Diese Graphik wurde mit RDF Gravity (RDF Graph Visualization Tool) erstellt:
<http://semweb.salzburgresearch.at/apps/rdf-gravity>

61 Das verkürzte RDF Dokument liegt unter <http://www.stud.fernuni-hagen.de/q6261035/Bachelorarbeit/PostItStamm.rdf> und wurde mit dem ICS FORTH VRP validiert:
<http://139.91.183.30:9090/RDF/VRP>

62 Im- und Export von Daten: <http://nulllogicone.net/RDF>

6.3 Web Services

Für den Austausch von Daten zwischen verschiedenen Anwendungen und OLI-it stehen Web Services⁶³ zur Verfügung. Stamm, Angler, PostIt, Code, TopLab und der Wortraum werden auf Anfrage im XML oder RDF Format ausgegeben und neue Elemente können in die Datenbank eingefügt werden.

Zudem gibt es einen Web Service, um zwischen Codes und Anglern zu vermitteln. Es besteht die Möglichkeit, einen Code gegen alle Angler, einen Angler gegen alle Codes oder einen Code gegen einen Angler abzugleichen.

Eine SOAP Nachricht als Anfrage an den Web Service, einen Code gegen einen Angler abzugleichen, enthält die GUID des Codes und die des Anglers:

```
POST /LogicService.asmx HTTP/1.1
Host: service.olii-it.com
Content-Type: text/xml; charset=utf-8
Content-Length: length
SOAPAction: "http://nulllogicone.net/schemata/CodeAnglerMatch"

<?xml version="1.0" encoding="utf-8"?>
<soap:Envelope
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xmlns:xsd="http://www.w3.org/2001/XMLSchema"
    xmlns:soap="http://schemas.xmlsoap.org/soap/envelope/">

    <soap:Body>
        <CodeAnglerMatch xmlns="http://nulllogicone.net/schemata/">
            <cguid>guid</cguid>
            <aguid>guid</aguid>
        </CodeAnglerMatch>
    </soap:Body>
</soap:Envelope>
```

Die Antwort auf eine solche Anfrage ist wieder eine SOAP Nachricht, die eine Boolesche Variable als Rückgabewert enthält:

63 OLI-it Web Services: <http://service.olii-it.com>

```

HTTP/1.1 200 OK
Content-Type: text/xml; charset=utf-8
Content-Length: length

<?xml version="1.0" encoding="utf-8"?>
<soap:Envelope
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xmlns:xsd="http://www.w3.org/2001/XMLSchema"
    xmlns:soap="http://schemas.xmlsoap.org/soap/envelope/">

    <soap:Body>
        <CodeAnglerMatchResponse xmlns="http://nulllogicone.net/schemata/">
            <CodeAnglerMatchResult>boolean</CodeAnglerMatchResult>
        </CodeAnglerMatchResponse>
    </soap:Body>

</soap:Envelope>

```

6.4 Verteilbarkeit

Verteilte Datenhaltung

6.5 OLIClient

Um Nachrichten zwischen Menschen zu vermitteln, wurde ein Client⁶⁴ als Referenz implementiert, der es dem Benutzer ermöglicht, Nachrichten zu erstellen, zu editieren, und nach Nachrichten zu suchen. Für den Austausch von Daten mit der zentralen Datenbank und für die Vermittlung werden Web Services in Anspruch genommen. Die Daten werden dezentral auf dem Rechner, auf dem der Client arbeitet, gehalten und können mit der zentralen Datenbasis synchronisiert werden.

6.6 UniSonus

Gerade Studenten der FernUniversität in Hagen erleben eine extreme Isolation im Studium. Dieser wird durch viele verschiedene Angebote der Universität, des AStA und der Studenten entgegengewirkt. Hierbei kommen alle oben beschriebenen Ansätze zur Nachrichtenvermittlung zum Einsatz.

Studenten an der FernUniversität Hagen können sich ein E-Mail Konto einrichten und mit bekannten Kommilitonen oder Ansprechpartnern der Universität privat kommunizieren.

64 OLIClient: <http://client.oli-it.com>

Eines der meistgenutzten Angebote für Fernstudenten sind die Newsgroups. Es gibt spezielle Newsgroups für die einzelnen Kurse, in denen inhaltlicher Austausch stattfindet, und allgemeine für den Gedankenaustausch und die Kontaktfindung zwischen Studenten.

Zudem wird ein IRC Chat angeboten, in dem offene Chats zu den einzelnen Fachbereichen stattfinden und Studenten eigene Channels bilden können, um sich in Echtzeit auszutauschen. Dort gibt es auch die Möglichkeit, einen privaten Chat einzurichten.

Die Virtuelle Universität bietet Webforen für einzelne Kurse. Der BSCW-Server (Basic Support for Cooperative Work) unterstützt das gemeinsame Arbeiten an Dokumenten, die Verabredung von Terminen und Online-Treffen.

Ein studentisches Angebot, die RuF Kontaktbörse, eröffnet den Studierenden der FernUniversität eine Möglichkeit anhand von Kriterien wie dem Fachbereich, den belegten Kursen, dem Hörerstatus, dem Bundesland sowie dem Postleitzahlbereich nach Kommilitonen zu suchen. Weitere studentische Kommunikationsangebote sind die Boards des AstA, die wie schwarze Bretter für Ankündigungen und Anzeigen dienen, und der Studentenserver, auf dem jeder Student seine eigenen Webseiten veröffentlichen kann.

Die oben erwähnten Angebote der FernUniversität haben gemeinsam, dass sie ausschließlich den Kommunikationsmöglichkeiten für und zwischen Studenten dienen und daher anmeldpflichtig und von der Außenwelt abgeschirmt sind. Weniger abgeschirmte Angebote sind ein von einem Studenten zur Verfügung gestelltes Wiki⁶⁵ und ein von einem Studenten betriebenes Forum⁶⁶.

Studenten finden anhand Kursen, Lokalität, Studiengängen: UniSonus⁶⁷ mit speziell angepasstem Wortraum, für jeden editierbar

65 Fernuni Wiki: <http://fernuni-wiki.de>

66 Studienservice: <http://www.studienservice.de>

67 UniSonus: <http://www.unisonus.net>

Literaturverzeichnis

- [AB04] ANTONIOU, G. und H. BOLEY: *Rules and Rule Markup Languages for the Semantic Web*. Springer-Verlag, 2004.
- [AH04] ANTONIOU, G. und F. VAN HARMELEN: *A Semantic Web Primer*. The MIT Press, Cambridge, 2004.
- [Bec06] BECK, K.: *Computervermittelte Kommunikation im Internet*. Oldenbourg, 2006.
- [BF00] BERNERS-LEE, T. und M. FISCHETTI: *Weaving the Web*. Texere, 2000.
- [BK03] BEIERLE, C. und G. KERN-ISBERNER: *Methoden wissensbasierter Systeme*. Vieweg, 2003.
- [BTF04] BUSSLER, C., V. TANNEN und I. FUNDULAKI: *Semantic Web and Databases*. Springer-Verlag, 2004.
- [DOS03] DACONTA, M., J. OBRST und T. SMITH: *The Semantic Web*. Wiley, 2003.
- [DGKM71] DROSDOWSKI, G., P. GREBE, R. KÖSTER und W. MÜLLER: *Der große Duden*. Fremdwörterbuch. Bibliographisches Institut AG, Mannheim, 1971.
- [FS05] FAGES, F. und S. SOLIMAN: *Principles and Practice of Semantic Web Reasoning*. Springer-Verlag, 2005.
- [FHLW03] FENSEL, D., J. HENDLER, H. LIEBERMAN und W. WAHLSTER: *Spinning the Semantic Web*. The MIT Press, Cambridge, 2003.
- [GP00] GOLDFARB, C. und P. PRESCOT: *Das XML-Handbuch*. Addison-Wesley, 2000.
- [GS02] GUMM, H. und M. SOMMER: *Einführung in die Informatik*. Oldenbourg Wissenschaftsverlag, 2002.
- [Hel96] HELBIG, H.: *Künstliche Intelligenz und automatische Wissensverarbeitung*. Verlag Technik, 1996.
- [Hel04] HELBIG, H.: *Grundlagen der Künstlichen Intelligenz*. FernUniversität in Hagen, 2004.
- [HILS03] HAAKE, J., C. ICKING, B. LANDGRAF und T. SCHÜMMER: *Verteilte Systeme*. FernUniversität in Hagen, 2003.
- [Min02] MINTERT, S.: *XML & Co*. Addison-Wesley, 2002.
- [MO96] MORRIS, M. und C. OGAN: *The Internet as Mass Medium*. Journal of Communication, 46. Jg., Nr.1, 1996.
- [Pow03] POWERS, S.: *Practical RDF*. O'Reilly, 2003.
- [SS04] STAAB, S. und R. STUDER: *Handbook on Ontologies*. Springer-Verlag, 2004.
- [SSU01] SCHWABE, G., N. STREITZ und R. UNLAND: *CSCW-Kompendium*. Springer-Verlag, 2001.
- [Woh03] WOHLFEIL, S.: *Sicherheit im Internet*. FernUniversität in Hagen, 2003.

Glossar

Algorithmus	Ein Algorithmus ist ein Verfahren zur Berechnung einer Funktion.
API	API steht für Application Programming Interface, also für eine Programmierschnittstelle, die Funktionsaufrufe auf der Programmierebene ermöglicht.
Axiom	Ein als wahr angenommener Grundsatz, der innerhalb eines Systems nicht begründet werden kann und muss.
B2B	B2B ist die Abkürzung für Business to Business und steht für Beziehungen zwischen Unternehmen.
Datalog	Datalog ist eine Datenbank-Programmiersprache, eine Untersprache der Hornlogik und die Schnittmenge von SQL und PROLOG.
Disjunktheit	In der Mengenlehre heißen zwei Mengen A und B disjunkt oder elementfremd, wenn sie kein gemeinsames Element besitzen.
Dublin Core	Der Dublin Core ist eine Grundmenge von beschreibenden Termen für die Kategorisierung von Webressourcen. Da der Dublin Core in RDF Dokumenten häufig verwendet wird, ist es sinnvoller, Terme daraus als selbst definierte zu verwenden.
Entität	Entitäten sind unterscheidbare, in der realen Welt eindeutig identifizierbare Objekte, die sowohl physischer als auch abstrakter Natur sein können.
Extension	Die Extension einer Klasse ist die Menge ihrer Instanzen.
GUID	Ein Globally Unique Identifier ist eine global eindeutige Zahl, die 128 Bit groß ist. Dadurch geht die Wahrscheinlichkeit, dass zwei gleiche GUIDs erzeugt werden, gegen Null. Wenn zusätzlich die global eindeutige MAC Adresse (Nummer der Netzwerkkarte) und der Zeitpunkt der Erstellung in die Berechnung einfließen, ist die Wahrscheinlichkeit gleich Null.
HTTP	Das Hypertext Transfer Protocol ist ein Protokoll zur Übertragung von Daten über ein Netzwerk. Es wird hauptsächlich eingesetzt, um Webseiten in Browser zu laden.
Inferenz	Inferenz ist der Oberbegriff für alle Formen des Schlussfolgerns.
Kardinalität	Kardinalität ist die Anzahl der Elemente endlicher Mengen.
Metamodellierung	Metamodellierung ist die Definition der Elemente des Vokabulars, die in einer Modellierungssprache verwendet werden. ⁶⁸
P2P	In Peer-to-Peer Netzwerken kommunizieren Gleiche mit Gleichen. Jedes Mitglied kann als Client und als Server fungieren.
Prolog	Prolog ist eine deklarative logische Programmiersprache. Ein Prolog Programm ist eine Sammlung von Hornklauseln, die zusammen eine Datenbasis, bestehend aus Fakten und Regeln, bilden. An diese Datenbasis können Anfragen gestellt werden.

68 Metamodeling Architecture of Web Ontology Languages:
<http://www.cs.man.ac.uk/~horrocks/Publications/download/2001/rdfsfa.pdf>

Semantik	Die Semantik einer Sprache bestimmt, auf welche Begriffe der zu repräsentierenden Welt sich die Sätze beziehen. Erst mit einer solchen semantischen Beziehung erlangen die Sätze der Repräsentationssprache eine Bedeutung, die zum Beispiel festlegt, ob ein Satz der Sprache in einer gegebenen Welt eine wahre Begebenheit bezeichnet oder nicht [BK03].
SQL	SQL, die Structured Query Language, ist eine deklarative Datenbanksprache für Relationale Datenbanken.
Syntax	Die Syntax einer Wissensrepräsentationssprache legt fest, wie die Sätze dieser Sprache aufgebaut sind [BK03].
Unicode	Unicode ist ein internationaler Standardzeichencode, der sich noch in der Entwicklung befindet. Langfristig wird für jedes Zeichen aller bekannten Schriftkulturen und Zeichensysteme ein digitaler Code festgelegt, wodurch das Problem verschiedener inkompatibler Kodierungen aus unterschiedlichen Ländern beseitigt werden soll.
Vererbung	Der Begriff Vererbung bedeutet in der objektorientierten Programmierung, dass von einer Klasse (Superklasse) weitere Klassen (Subklassen) abgeleitet werden können, die die Eigenschaften und Methoden der Superklasse erben. Eine Subklasse verfügt meist über weitere Eigenschaften und Methoden, so dass sie als Spezialisierung der Superklasse betrachtet werden kann.
Vokabular	Das Vokabular ist eine Menge aus URI references.

Stichwortverzeichnis

Chat	6	Vokabular	34
digitale Signatur	40	OWL Service	32
Dublin Core	23	Public Key Verschlüsselung	41
E-Mail	1	RDF	22
Gehirn		Datenmodell	23
Neuron	43	Graph	23
Neurotransmitter	44	Inferenz	29
Synapse	44	RDF Schema	25
Global Brain	42	RDF/XML	24
Interpretation	29	Reification	24
Kreislauf		Ressource	23
Angler	49	Semantik	29
Code	49	Tripel	23
PostIt	47	Regeln	38
Stamm	47	RSS	25
TopLab	47	RuleML	39
Logik	38	Semantic Web	18
Hornlogik	39	Softwareagent	18
Prädikatenlogik	38	SPARQL	31
Metadaten	19	SWRL	39
Modelltheorie	29	Tagging	33
Namespace	22	Taxonomie	33
OLI-it	47	URI	22
Abgleich	52	Usenet	2
Kreislauf	47	Vertrauen	40
Realisierung	52	Web Service	20
Vermittlungslogik	49	Webforen	7
Wortraum	48	Weblogs	9
Ontologie	34	Wikis	8
Open-World-Assumption	37	Wissensrepräsentation	38
OWL	32	Wortraum	
Description Logic	38	Baum	48
Full	38	Knoten	48
Lite	37	Netz	48
Mapping	37	Zweig	48
Restriktionen	35	XML	20

Abbildungsverzeichnis

Abbildung 1: Die Schichten des Semantic Web	20
Abbildung 2: Ein einfaches RDF Statement als Graph dargestellt	23
Abbildung 3: RDF Schema	26
Abbildung 4: Ein einfaches RDF Schema	27
Abbildung 5: Verwendung von RDF Schema	28
Abbildung 6: Eine einfache Restriktion	36
Abbildung 7: Verbindung zweier Neuronen	43
Abbildung 8: Schematische Darstellung einer Synapse	44
Abbildung 9: Der Kreislauf von OLI-it	47
Abbildung 10: Elemente des Wortraums	48
Abbildung 11: Ein Code mit passendem Angler	50
Abbildung 12: Beziehungen zwischen Klassen	58
Abbildung 13: Werte für eine Markierung	59
Abbildung 14: Ein PostIt mit seinem Stamm	60

Erklärung

Hiermit versichere ich, die vorliegende Arbeit selbstständig und unter ausschließlicher Verwendung der angegebenen Quellen und Hilfsmittel verfasst sowie Zitate kenntlich gemacht zu haben.

München, Mai 2006

Kathrin Dentler