

Reflexion zur VU

Vergleich der paradigmatischen und technischen Konzepte des „Semantic-Web“ mit anderen Konzepten der Datenverarbeitung

Was sind die konzeptuellen Unterschiede von
„Semantic-Web“ zu SOA und Datenbanken?

Wolfgang Friedhuber

Mat.Nr.:7430***

Stud.: B 033/645

Graz: 5.4.2011

Korrigierte und überarbeitete Version vom 18.7.2011

SE 521.133

Konzepte und Technologien für eine semantische
Datenverarbeitung

Leitung: Dr. Patrick Sahle

Inhaltsverzeichnis

1	Einleitung	4
2	Abriss der historischen Entwicklung	8
2.1	Die „Steinzeit“	8
2.2	Das „Mittelalter“	9
2.3	Die „Neuzeit“ der Computer	12
3	Die Auswirkungen	14
3.1	Pluralisierung der Softwareerstellung	14
3.2	Gewichtungsänderung technischer Parameter	15
3.3	Komplexitätsreduktion	16
3.4	Paradigmen des WWW	17
4	Begriffsbildung und Bedeutung	19
4.1	Allgemeine Betrachtungen	19
4.2	Der Begriff „Semantic Web“	22
4.3	Was ist mit „Semantik“ gemeint	27
4.4	Schlussfolgerung	33
5	Semantic Web Konzepte	35
5.1	Die Basiskonzepte	36
5.1.1	Die Architektur	37
5.1.2	Die semantischen Konzepte	40
5.2	Die Notationen	42
6	Die Technik	48
7	Vergleich von Realisierungsalternativen	52
7.1	Anwenderanforderungen	52
7.2	Lösungsansätze	54
7.2.1	Konventionelle Lösung	54
7.2.2	Lösung als SOA (CORBA)	57
8	Resümee	66

Abbildungsverzeichnis

2.1	OSI 7 Schichtenmodell	12
3.1	Client-Server-Paradigmenwellen	14
4.1	Semantic Web im Internet	26
5.1	Die 4 Versionen der Referenzarchitektur	38
5.2	CFL-Architekturmodell von Gerber	39
5.3	Zuordnung Berners-Lee V4 Modell zu CFL-Modell	40
6.1	Kommunikationsequenz: Browser ist Endpunkt der Kommunikation	48
6.2	Architektur eines Browser unter Windows	49
6.3	Anwendungsadressierung der TCP/IP Kommunikation	50
6.4	Schema einer http-Sitzung	51
7.1	Anwendungsfälle für Semantic Web	61
7.2	Unterschied Relationale DB vs. Ontologie	62
7.3	Oracle SOA-Suite	62
7.4	SOA - Entwurfsparadigmen	63
7.5	Kommunikationsbeziehungen SOA-Komponenten	63
7.6	Aspekte einer SOA	64
7.7	Schichtenmodell der Dienste	64
7.8	SOA Services-Modell	65

1 Einleitung

Das Semantic Web wird [...] das bestehende WWW erweitern und für den Benutzer stark vereinfachte Dienste im Bereich der Informationssuche, des E-Business, des E-Government, des E-Recruiting und des Wissensmanagements anbieten. Branchen-Ontologien werden künftig die wesentlichen Begriffe und Beziehungen für eine bestimmte Branche definieren, so dass eine Interoperabilität von Webdiensten erreicht wird. Metadaten werden die Transformation digitaler Inhalte für unterschiedliche Ausgabegeräte und Nutzergruppen unterstützen.¹

So schreiben Studer und Wahlster 2007 im Vorwort des Buches *Semantic Web*. Berners-Lee als „Erfinder“ des WWW und nun auch Erfinder von „Semantic Web“ selbst schreibt:

The Semantic Web is not a separate Web but an extension of the current one [...].²

Diese Erweiterung soll den für Menschen gedachten „Content“³ des WWW⁴ für Computer „verständlich“⁵ machen. So oder so ähnlich lauten die Erklärungen für

¹Hitzler et al., 2008, Vorwort.

²Berners-Lee/Hendler/Lassila, 2001.

³Content = Inhalt; Gemeint ist hier der Inhalt der über das WEB gesendeten Information. Der Begriff ist als Teil der webspezifischen Ausdrucksweise zu sehen.

⁴World Wide Web: Bezeichnung für das auf tcp/ip (Transmission Control Protocol/Internet Protocol; Kommunikationsnormen welche die Netzwerkadressierung und den Informationstransport spezifizieren) und http (Hypertext Transfer Protocol; Kommunikationsnorm, welche die Codierung der transportierten Daten festlegt) basierende weltweite Kommunikationsnetz

⁵Ob hier der Begriff „verständlich“ gerechtfertigt ist oder nicht, ist diskussionswürdig. Es kann der Standpunkt vertreten werden, das „Verstehen“ eine spezifisch menschliche Eigenschaft und damit nur auf die zwischenmenschliche Kommunikation anzuwenden ist. Für Maschinen wäre dann der Begriff „verarbeitbar“ adäquater. Allerdings ist gerade im Semantik-WEB das Ziel, die Kommunikation in ihrer inhaltlichen Bedeutung für Maschinen erfassbar zu machen und Schlussfolgerungen aus Bedeutungen zu ziehen. Diese Fähigkeit wird mit „Verstehen“ treffender ausgedrückt.

Semantik-WEB.

Was diesen Erklärungen anhaftet, ist ihre unpräzise, teilweise verwirrende Verwendung von Begriffen. Was meint Studer mit „Interoperabilität“?⁶ Die „Interoperabilität von Webdiensten“⁷ ist doch das „Um und Auf“ von Webkommunikation. Die Unterstützung der Inhalte für verschiedene Ausgabegeräte durch Metadaten gibt es doch schon seit es Computer gibt – wie sonst könnte ein Drucker Texte ausgeben?⁸ Interoperabilität als Protokollnormierung des Datenaustauschs zwischen Kommunikationspartnern kann hier doch kaum gemeint sein, die gibt es ja bereits seit es Kommunikation gibt. Meint Studer hier etwa zentrale Dienstbeschreibungen wie sie von SOA-Konzepten⁹ her vorgeschlagen wurden? Soll nur auf die triviale Tatsache hingewiesen werden, dass, wenn zwei kommunizieren, die sich verstehen müssen? Oder meint Studer die zur Zeit noch inkompatiblen Webdienste, für die nun unter dem Titel „Semantic-WEB“ Zug um Zug Schnittstellenadapter angefertigt werden solle? Besonders verwirrend ist der Hinweis auf die Branchen im Zusammenhang mit Ontologien.¹⁰ Das lässt wieder vermuten, dass die Interoperabilität eventuell nicht auf der Ebene von technischen Dienstschnittstellen liegt sondern eher auf der Ebene von Übersetzungsleistungen.¹¹

⁶Es wird bei Hitzler klarer, auf was sich „Interoperabilität“ bezieht (vgl. Hitzler et al., 2008, S. 11). Ich möchte den Begriff „Interoperabilität“ hier aber doch als Beispiel problematisieren, da er zeigt, wie mit Begriffen gearbeitet wird. Es werden scheinbar Konzepte und Eigenschaften konkret genannt. Geht man dem aber dem nach, so stellt sich oft heraus, dass eigentlich die breiteste der Deutungsmöglichkeiten zutrifft - also genau genommen nichts konkretes ausgesagt wird. Auch hier: Die „Interoperabilität“ meint hier den Datenaustausch zwischen unterschiedlichen Anwendungen - nicht die Netzkommunikation oder die Kommunikation eines Clients mit seinem Service. Die Interoperabilität von Diensten klingt irgendwie konkret - sagt aber nichts aus, da Dienste (Services) untereinander nicht kommunizieren. Sie kommunizieren mit Dienstanwender (Clients). Was hier eigentlich gemeint ist, ist die Tatsache, dass die Daten der einzelnen Dienste nicht von anderen Diensten verarbeitet sind (so wird z.B. eine Mailnachricht für einen Browser nicht interpretierbar sein).

⁷a. a. O., Vorwort.

⁸Auch die Umsetzung von proprietären Metadaten durch spezifische Treiber oder Adapter ist seit langem Stand der Technik.

⁹SOA = Service Oriented Architecture. In SOA wird ein Dienst-Broker (Dienstanbieter) als zentrales Dienstverzeichnis vorgeschlagen. In diesem Dienstverzeichnis kann auch die Schnittstelle beschrieben sein. Interoperabilität ist dann dadurch gewährleistet, dass

1. Der Dienst im Dienstverzeichnis gelistet ist.
2. Die Schnittstelle spezifiziert ist.

(Vgl. Sayegh, 1997, S.41-45 u. S. 71-78), (sowie a. a. O., S. 79-88), (auch Dostal et al., 2005, S. 27-36).

¹⁰Es soll hier die Problematik der möglichen Zugänge pointiert dargestellt werden. Die genannten Themen werden im Zuge dieser Arbeit einer Klärung zugeführt werden.

¹¹Etwa in der Art, von Adapter für Datenbankschemata oder XML-DB-Frontends (vgl. Türker/Saake, 2006, S. 467-526)

Berners-Lee prägte den Begriff „Semantic Web“. Bei seiner Darstellung¹² von Semantic Web handelt es sich um eine Beschreibung, die anno 1960 in Science Fiction Romanen auch verwendet wurde. Es ist dabei klar, dass es sich um eine Vision handelt, die keinerlei technische Information beinhaltet. Die Darstellung ist auf der Stufe von „wir werden mit Robotern reden können, die werden die Arbeiten für uns erledigen“ - wie man eben bis in die 1970er Jahre gemeint hat, dass sich die Computertechnik entwickeln würde. Eine Klärung, was real unter „Semantic Web“ beschrieben wird, erfolgt nicht.

Was ist da eigentlich los? Ist „Semantic Web“ wieder eine der „Marketingwellen“, wie sie in der Software so häufig sind? Schon die modulare Softwareerstellung, danach die Objektorientierung und danach die komponentenbasierte Software-Erstellung versprochen das Ende des Software-Problems und mit MDA¹³, SOA¹⁴ und ähnlichen „Paradigmen“ sollte die Software endlich die Anwenderbedürfnisse wirklich befriedigen.

Diese Hypes, die in Wellen auftreten, sind das nur marktschreierische Produkpropagierungen oder tatsächlich technische Konzepte – oder sogar mehr? Was ist mit den Behauptungen zu den Leistungen von WWW, WEB 2.0 und WEB 3.0? Was bedeuten die Begriffe, die kaum mehr Rückschlüsse auf Konzepte zulassen?¹⁵ Sind das tatsächlich so gravierende Änderungen, dass die Bedeutung über lokale Technikanwendungen hinaus geht und sie auch gesellschaftlich und weltbildverändernd wirken? Ähnliche Fragen stellen sich auch beim Thema „Semantic Web“¹⁶: Künstliche Intelligenz, Datenbankabfragen und Datamining, Vernetzung und das Anbieten von Diensten gibt es doch schon lange. Was verbirgt sich also hinter diesem Schlagwort „Semantic Web“?

Um es gleich vorweg zu nehmen: hinter Akronymen wie WWW, WEB2.0 und auch dem Schlagwort „Semantic Web“, manchmal auch als „WEB 3.0“ bezeichnet,¹⁷ verbirgt sich tatsächlich ein Paradigmenwechsel der den Namen wirklich verdient. Worin dieser Wechsel besteht, was er für Folgen nach sich zieht und wo

¹²Siehe Berners-Lee/Hendler/Lassila, 2001.

¹³MDA = Modell Driven Architektur

¹⁴SOA = Service Oriented Architecture

¹⁵Die Reflexionen zur Begriffsbildung siehe Kap. 4 S. 19.

¹⁶Zur Reflexion des Begriffes „Semantic Web“ siehe Kap. 4.2 S. 22.

¹⁷Vgl. Einleitung Geisler, 2009, S. 11.

die Grenzen und Probleme liegen, versuche ich im Nachfolgenden zu analysieren und im Vergleich zu früheren technischen Paradigmen darzustellen.

Ich muss gestehen, dass ich sehr ernsthafte Probleme habe, in diese moderne Gedankenwelt von Werbetrommlern, Visionen, technische Aussagen, Selbstdarstellungen und Community-Slang – alles bunt durcheinandergemixt – einzudringen. Es geht mir daher im Weiteren nicht um eine exakte und vollständige Beschreibung der Computerentwicklung oder des Entwicklungsstandes des „Semenatic Web“, sondern um den Versuch, den roten Faden der technischen Realität zu finden und aufzunehmen.

Zur Entschuldigung meines Unvermögens das Gebiet geistig rasch und korrekt zu erfassen, möchte ich anführen, dass es anscheinend den Schreibern der Erklärungsbücher nicht anders geht. An Wiederholungen, Tautologien und Widersprüchlichkeiten in den Texten glaube ich zu erkennen, dass die korrekte Erfassung nicht einfach – oder gar unmöglich ist. Das Problem der Erfassung liegt nicht darin begründet, dass es um Zukunftsentwürfe geht. Die ließen sich relativ widerspruchsfrei darstellen, das Problem liegt in der unscharfen, offenen Begrifflichkeit und ihrer bedarfsspezifischen Interpretationen. Das ist ein Kennzeichen der neuen Paradigmen: Sie sind „unfassliche“, offene Konzepte.

2 Abriss der historischen Entwicklung

2.1 Die „Steinzeit“

Als computertechnische „Steinzeit“ soll hier der Zeitraum bis zum Aufkommen der IBM-PC¹⁸ und dem Markteintritt von Microsoft, vor allem mit „Windows“ als PC-Betriebssystem, gesehen werden. In dieser „Steinzeit“ galten für Computer Paradigmen die aus der allgemeinen Technik und der mathematisch basierten Informationstechnik stammten.

Bis in die 1980er Jahre waren Rechenmaschinen hauptsächlich großen Institutionen vorbehalten. Um 1980 herum, begannen Firmen wie ATARI Spielkonsolen zu vertreiben. APPLE eröffnete mit seinem legendären APPLE II¹⁹ das Zeitalter des „Personal Computers“ – also des Computers für jedermann.²⁰ In Deutschland und Österreich war es vor allem der „Comodore 64“²¹. Der C64 der Firma Comodore führte, aufgrund seines niedrigen Preises, zu einer massenhaften privaten Nutzung von Computern.

Diese Welle der Nutzung war durch folgende Merkmale gekennzeichnet:

- möglichst geringe Kosten
- möglichst einfache Bedienung / Bedienerführung
- Performance der Software orientiert sich an Hardwareperformance

¹⁸Genauer: Bis zum Aufkommen der IBM Hardwarearchitekturen mit den Typen „PC XT“ und „AT“ um ca. 1985 (vgl. Meinhardt, 1995, S. 24-25).

¹⁹Vermarktungsbeginn ca. 1977

²⁰Mit dem Apple II kündigte sich, durch seine Modularisierung der Hardware, ein Paradigmenwechsel an.

²¹Markteinführung ca. 1982. Das Vorgängermodell „VC 20“ aufgrund seines niedrigen Preises der erste wirklich für die Massen leistbare Computer. Brauchbare Ressourcen bot aber erst der „Comodore 64“ (C64).

- Schwerpunkt der Anwendungen für PC sind Spiele und Programmerstellung
- Stand-Alone-Betrieb (unverbundener Betrieb)
- möglichst geringer Ressourcenverbrauch
- Programmerstellung durch den Anwender (User)

Natürlich gab es bald Modems, mit denen PC-User andere Computerbesitzer anrufen konnten um Programme auszutauschen. Das änderte aber nichts an der Anwendung des Rechners als lokale Maschine.

Die Bedienung wurde durch ein kleines Betriebssystem mit sehr geringem Leistungsumfang gewährleistet. Die Programmiersprache BASIC, die als Interpreter oder halbcompilierende Interpreter bei fast allen Personal Computer (PC) Standard war, sollte einfache Programmerstellung ermöglichen. Vor allem BASIC hatte das Ziel, die Maschinennutzung jeden möglichst in Umgangssprache²² zu ermöglichen. Das Leitparadigma blieb aber der technische Aspekt.

Bei Institutionen wie Universitäten, Firmen usw., also im „professionellen Bereich“ wurden Computer verwendet, die mit vollwertigen Betriebssystemen wie UNIX arbeiteten und standardmäßig in Client-Server-Netzen liefen. Aber für die große Masse der Anwender galt:

- Einzelplatzbetrieb
- optimierter Ressourcenverbrauch
- Anwendungsschwerpunkt: Spiele und eine techniklastige Nutzung (analog zu CB-Funker-Szene)

2.2 Das „Mittelalter“

Mit der Modularisierung der Computerkomponenten, die Apple begann und IBM konsequent fortsetzte, sanken die Preise bei stetig steigender Leistungsfähigkeit. Das Konzept, die Hardwarearchitektur in genormte Einheiten zu teilen, die über

²²Englisch ist hier gemeint.

Busse verbunden waren, ermöglichten eine Verbreiterung der Zulieferbasis und eine Spezialisierung der Lieferfirmen und führten im Weiteren zu einer Optimierung der Komponente. An der Verwendung als Einzelplatzmaschine änderte das wenig. Neben den Spielen wurden aber jetzt vermehrt Anwendungen für den Nutzer möglich. Die Nennung von „Word-Star“²³ soll diese Phase charakterisieren. Ansonsten stagnierte die Entwicklung eher. Die Bestrebungen Microsofts, aus dem PC und seinen Teilen möglichst hohen Gewinn durch Massenanwendungen zu schlagen förderte einerseits die Massenanwendung änderte aber den Fokus der Anwender. So versuchte Microsoft von Anfang an proprietär zu sein. Nichts von MS (Microsoft) sollte zu Bestehendem passen. Alles sollte von MS bezogen werden müssen. Um das auch gegen rationale Überlegungen der User zu erreichen, begann die Zeit der Hypes. Die technische Sinnhaftigkeit trat hinter griffigen Schlagworten für die Vermarktung zurück. Der erste Paradigmenwechsel war geboren.

Folgende Schwerpunkte änderten sich:

- Ressourcenverbrauch wurde zunehmend unbedeutend
- DV-technisches Verständnis wurde zurückgedrängt (Icon-Bedienung)
- Vermarktung wurde zentrales Entwurfskriterium
- Anwendungen wurden Hauptfokus
- Entwicklungsanforderungen wurden auf die Hardwareseite verlagert
- Performance löst sich von den physikalischen/physiologischen Prämissen

Die Argumentationsketten wurde praktisch umgedreht: War vorher die Begründung, etwas sei softwaretechnisch zu lösen, da es leichter änderbar und daher kostengünstiger war, begann nun die Zeit, wo die Softwarekosten die Hardwarekosten überstiegen. Die Software lagerte ihre Performance- und Ressourcenprobleme in die Hardware aus. Sie, vor allem die von Microsoft, gab der Hardware vor, was sie zu leisten hatte.

Das führte zu einer nahezu explosiven Entwicklung auf dem Hardwaresektor, die Leistung und den Preis betreffend. Microsoft (MS) schaffte es, dass Begriffe wie

²³Ein Textverarbeitungsprogramm aus den 1980er Jahren.

Effizienz, Performance, Fehlerfreiheit, Rationalität und Architektur auf dem Softwaresektor völlig verändert wurden; auch im professionellen Anwendungsbereich. Selbst Steuerungen in kritischen Produktionsstätten begannen „MS-Windows“ zu verwenden, obwohl bekannter Maßen pro Tag mehrmals gebootet²⁴ werden musste. Ein Tatbestand der vor „Windows“ unakzeptabel gewesen wäre.

Gleichzeitig, obwohl – oder gerade weil – der Zugang zum PC proprietärer,²⁵ also herstellerspezifischer wurde, verbreiteten sich die MS Produkte rasant – auch im institutionellen universitären Bereich. Dadurch wurden, ausgehend von den Universitäten und Forschungsstätten, die immer schon vernetzte Rechner hatten, nun auch PCs immer mehr vernetzt. Waren die vernetzten Rechner bis dahin wenige, unerschwingliche Großrechner, so waren es jetzt relativ preiswerte PCs in großer Stückzahl. Auch die vielen PCs der „normalen PC-User“ zu vernetzen, war der nächste logische Schritt. Die Basis für diesen Schritt wurde durch die Versuche der Schnittstellenstandardisierung gelegt. So versuchte die ISO²⁶ durch ihr 7-Schichtenmodell eine verbindliche Norm für Netzwerkkommunikation einzuführen.²⁷ Durch das OSI-Referenzmodell setzte sich der Applikationsarchitektur in Schichten durch. Diese Schichtenarchitektur ist die Basis für die Erstellung von anwendungsspezifischen Softwaremodulen und leitete die komponentenbasierte Softwareentwicklung ein.

Als entscheidender Paradigmenwechsel kann hier festgehalten werden:

- Abkehr von ressourcendominierten Denken (Effizienz)
- Abkehr von technischer Sichtweise der Datenverarbeitung (Marktmechanismen wurden wichtiger)
- Normierung der Kommunikationsschnittstellen

²⁴Also die Maschine heruntergefahren und neu gestartet werden musste. Nicht freigegebener dynamischer Speicher führte zu Systemabstürzen und „hängenden“ Systemen bei längerer Nutzungszeit.

²⁵Zu der Zeit war eigentlich der Zugang zum PC immer proprietär. Während aber Apple und auch andere Hersteller umfassende Dokumentation lieferten, um dem Anwender Internas zugänglich und verständlich zu machen, vertrat Microsoft von Anfang an eine andere Strategie. Das ist hier mit „proprietär“ gemeint.

²⁶ISO = International Organization for Standardization. Eine Institution, die versucht internationale Normierungen voranzutreiben.

²⁷Siehe OSI-Referenzmodell; OSI = Open Systems Interconnection Reference Model (siehe auch Fig. 2.1 S. 12). Auch wenn gerade im Internetsektor sich anstelle des OSI-Modelles TCP/IP durchsetzte brachten die Standardisierungsversuche doch eine Reduktion proprietärer Kommunikationslösungen mit sich und ermöglichten so den Schritt zur offenen Massenkommunikation.

Internet-Protokolle								
OSI-Schicht		Internet-Protokoll-Suite						Internet-Schicht
7	Anwendung	File Transfer	Electronic Mail	Terminal Emulation	Usenet News	Domain Name Service	World Wide Web	Art der Kommunikation
6	Darstellung	File Transfer Protocol (FTP) RFC 959	Simple Mail Transfer Protocol (SMTP) RFC 821	Telnet Protocol (Telnet) RFC 854	Usenet News Transfer Protocol (NNTP) RFC 977	Domain Name Service (DNS) RFC 1034	World Wide Web (WWW) RFC	Anwendung Applikation
5	Sitzung							
4	Transport	Transmission Control Protocol (TCP) RFC 793					User Datagram Protocol (UDP) RFC 768	Transport Host to Host Kommunikation
3	Netzwerk	Address Resolution Protocol (ARP) RFC 826	Internet Protocol (IP) RFC 791				Internet Control Message Protocol RFC 792	Internet
2	Sicherung	Ethernet		Token Ring	DQDB	FDDI	ATM	lokales Netzwerk
1	physikalische Übertragung	Twisted Pair	Lichtwellenleiter	Coaxkabel		Funk	Laser	Netzzugriff

Bild 1: Kommunikationsprotokolle im Internet

Abbildung 2.1: OSI 7 Schichtenmodell

(Quelle Gretzinger, 2007, S. 482)

- Einführung geschichteter Softwarearchitektur
- Vernetzung der Computer (zuerst noch ohne großen Vorteil)
- Einführung fehlerbehafteter Systeme (Fehlertoleranz der Anwender)

Die Einführung fehlerbehafteter Systeme ist deshalb genannt, da es eine wesentliche Eigenschaft des WWW ist, fehlertolerant zu arbeiten.

2.3 Die „Neuzeit“ der Computer

Das Fallenlassen des Paradigmas der Effizienz führte zu einer Leistungssteigerung der Hardware die vorher unvorstellbar war. Die Abkehr von der technikbasierten Rationalität ermöglichte Konzepte, die auf neuen Entwurfskriterien basierten. Die Akzeptanz von fehlerhaften Lösungen erlaubte die protoypenhafte Erprobung in Echtumgebungen und ermöglichte so Entwicklungen, die vorher nicht denkbar waren. Die Vernetzung stellte neue Einnahmequellen, neue Lösungsmöglichkeiten und neue Anwendungsmöglichkeiten in Aussicht. Der zweite fundamentale Para-

digmenwechsel kündigte sich an: Das WWW.²⁸

Dieser Paradigmenwechsel bedingte:

- Plattformunabhängige textbasierte Kommunikationsprotokolle
- Betriebssystemerweiterung für Netzkommunikation (Browser)
- Freigabe der Telefonnetze zur Datenübertragung
- Bildung von weltweiten Communities außerhalb von Institutionen

Das fundamental Neue war die Lösung der textbasierten Kommunikationsprotokolle. Die eigentlichen Daten (Texte) und der Steuercode für die Darstellung wurden gemeinsam dargestellt.²⁹ Verarbeitet wurde der Datenstrom auf allen Maschinen – unabhängig vom verwendeten Betriebssystem – durch Browser als Netzwerkendpunkt. Diese Browser hatten die Klartextsteuercodes zu erkennen und darzustellen.

²⁸Zu den Wellen in der Kommunikationsprinzipien siehe auch Fig. 3.1 S. 14.

²⁹Diese gemeinsame Betrachtung von Daten und Code entspricht auch dem objektorientierten Paradigma.

3 Die Auswirkungen

3.1 Pluralisierung der Softwareerstellung

Durch diese Konzepte der offenen Schnittstellen und der vernetzten Kommunikation kam aber auch ein anarchisch/demokratischer Ansatz in das Lösungskonzept.³⁰ Die Web-Daten waren in ihrer Gesamtheit als Text lesbar und konnten in jedem Texteditor erstellt und bearbeitet werden. Der textbasierte Kommunikationsansatz erschwert es zentralen Stellen, proprietäre Protokolle einzuführen. Der Preis dafür war eine Vervielfachung der Netzlast – also ein Verlust an Effizienz.

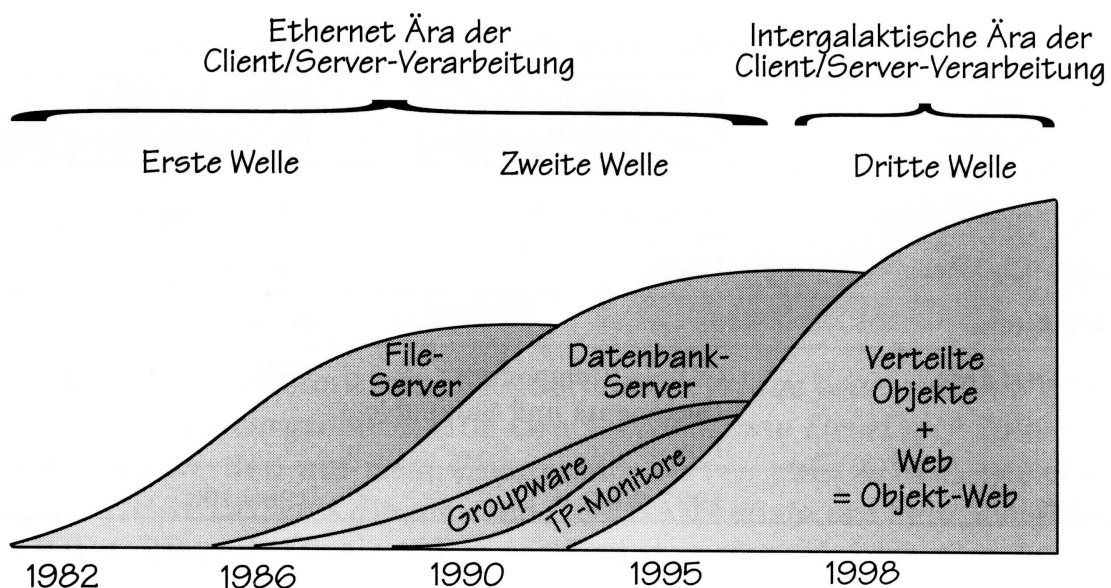


Bild 2.4. Die Client/Server-Wellen

Abbildung 3.1: Client-Server-Paradigmenwellen

(Quelle Orfali/Harkey/Edwards, 1998, S. 51)

³⁰Diese anarchische Komponente war von Berners-Lee durchaus beabsichtigt.

3.2 Gewichtungsänderung technischer Parameter

An der Veränderung der Kriterien der technischen Rahmenbedingungen der Kommunikationsprotokolle ist zu sehen, worin der Paradigmenwechsel bestand: Ein früherer Lösungsansatz wäre gewesen, dass man Kommunikationsnormen einführt, an die sich Sender- und Empfängeranwendung zu halten haben. Diese Norm hätte auf Effizienz in der Übertragung und der allgemeinen effizienten Ressourcennutzung bestanden.³¹ Für unterschiedliche Anforderungen wären eben unterschiedliche Protokollnormen verwendet worden. Im professionellen Bereich ist das auch heute noch so.³² Vor allem bei Zertifikationen muss immer noch weitestgehend nach klassischem Paradigma vorgegangen werden.³³ Im professionellen Bereich ist es auch heute noch notwendig, Effizienz, Performance und klare Dokumentation zu berücksichtigen. Proprietäre Lösungen sind hier vielfach noch anzutreffen.³⁴ Aber auch hier wird zunehmend versucht, den Kostenvorteil zu nutzen, der durch die WEB-Ideologie³⁵ möglich wird und die älteren Lösungen und Konzepte zu ersetzen.³⁶

³¹Wie gravierend diese Änderung in ihrer Auswirkung war, möchte ich an einer kleinen Anekdote aus der Zeit der Telefonfernwartung demonstrieren: War ca. um 1995 eine per Fernkorrektur eingespielte Fehlerkorrektur in die Fernmeldevermittlung HICOM noch 10-50 Bytes groß, so war eine ähnliche Korrektur ca. um 2005 in die HiPath im Megabyte-Bereich. Aufgrund der gesunken Leitungskosten wurden nämlich nicht wie vorher nur die geänderten Befehle ausgetauscht, sondern ganze Anwendungsteile durch eine Korrekturversion ersetzt.

³²Siehe z.B. Krankenhausinformationssysteme (KIS; englisch: HIS = Hospital Information System) oder in Wartungsnetzwerken wie SIRA (Secured Infrastructure for Remote Access), aber auch CORBA (Common Object Request Broker Architectur) sind noch weitgehend nach den alten Paradigmen entstanden. CORBA war aber bereits der Versuch, die neuen Paradigmen industriell einzubringen.

³³Als Beispiel für die Anforderungen bei Zertifikationen (siehe Sutter, 2007).

³⁴Vgl. Stohlmann, 2006.

³⁵Unter „WEB-Ideologie“ will ich hier das Konzept von generischen, offenen und meist kostenfreien Informations-technologiekomponenten verstehen.

³⁶Noch in meiner Tätigkeitszeit bei SIEMENS wurde daran gearbeitet, das Servicenetzwerk SIRA durch Standardkonzepte wie *Questa-Remote Services* und *Hewlett-Packard Remote Solution Standards* zu ersetzen. Diese Services sind zwar auch im hohen Maße spezifisch (funktionsspezifisch) – setzen aber vermehrt auf das SOA-Konzept und Standardisierung. Zum Zeitpunkt meines Ausscheidens aus Siemens im Jahre 2008, war es noch nicht gelungen, mit diesen standardisierten Alternativen die Funktionalität des proprietären SIRA-Konzeptes abzudecken. Einer der Gründe für die Probleme war die differenzierte Servicelandschaft mit ihrer Produktvielfalt. Für „Semantic Web“ Realisationen dürfte in Punkto Komplexität hier eine Parallelität gegeben sein.

3.3 Komplexitätsreduktion

Man könnte auch sagen, mit dem WWW ist der erste wirklich fundamentale Paradigmenwechsel in der datentechnischen Welt erfolgt, der die vorherigen Paradigmen zusammenfasst und zu einem unumkehrbaren Ereignis machte: Die Art, wie in der EDV mit Komplexität umgegangen wird, ist fundamental anders geworden. Durch die eingangs aufgezeigte Entwicklung hin zur modularen und komponentenbasierten Entwicklung, wurde es möglich, große Softwareprojekte in Teile zu brechen und iterativ zu entwickeln. Vor allem im Zusammenspiel mit der objektorientierten Softwarearchitektur und durch Einsatz von WEB-Services wurde es möglich, diese Teile verteilt zu realisieren.

Die Antwort auf Komplexität ist heute eine Modularisierung, die auf der Basis von möglichst einfachen Protokollen [...] eine verteilte Bearbeitung erlaubt. An die Stelle großer singulärer Lösungen treten jetzt Netzwerke kleiner Bausteinlösungen.³⁷

Wurde früher versucht, Komplexität durch umfassende Modellbildung und durch Algorithmen abzubilden und so beherrschbar zu machen (neuronale Netze, DB-Schemata, Programmialgorithmen usw.), so zielt das WWW-Paradigma auf Komplexitätsreduktion an sich!³⁸ Das Summenproblem wird weder dargestellt, noch analysiert oder modelliert. Es wird in einfache Teilprozesse aufgelöst. Damit werden keine umfassenden Lösungen angeboten, sondern komplexitätsreduzierte Teilmodelle spezifiziert, auf deren Basis jeder seine Lösungen ausprobieren kann (evolutionär entwickelte Lösungen in offenen Systemen). Das komplexe Gesamtziel ist nur mehr durch eine Vision vertreten, die aber nicht technisch spezifiziert wird.³⁹ Das ursprüngliche komplexe Summensystem soll sich evolutionär aus der Integration der unterschiedlichen Teillösungen bilden.⁴⁰

³⁷Patrik Sahle.

³⁸„Allerdings muss man beachten, dass die Komplexitätsreduktion auf den Basisebenen erst den Aufbau neuer Komplexitäten auf den Anwendungsebenen fördert.“(Patrik Sahle)

³⁹Gerade diese fehlende Spezifikation des Gesamtzieles ist es, was das eingangs erwähnte schwierige Eindringen in die Thematik verursacht. Da keine Spezifikation angefertigt wird, kann auch kein Begriffsschema aufgebaut werden. Summenmodell und Begriffswelt entstehen erst bei der modularen Realisierung. Aufgrund der verteilten Realisierung entstehen die Modelle und Begriffe unkoordiniert, redundant und uneinheitlich, jeweils nur mit einer modulspezifischen Gültigkeit.

⁴⁰Die radikale Form dieses Paradigmas ist nur in der WEB-Community vertreten. Die Softwareindustrie und teilweise auch universitäre Einrichtungen und Institute arbeiten meist noch näher am klassischen Vorgehen von

Nun weiß man aus der Komplexitätsforschung (im IT-Bereich aber auch aus den philosophischen Überlegungen) dass dieser Ansatz zwei Konsequenzen hat:

1. Er wirkt informationsvermehrend
2. Die volle Komplexität kann damit nicht abgebildet werden⁴¹

3.4 Paradigmen des WWW

Das Paradigma des WWW nimmt beide Konsequenzen in Kauf und macht sie zur Entwurfsrichtlinie. Damit ist eine fundamentale Abkehr vom technischen Paradigma der theoretischen Problemerkennung und dessen möglichst vollständigen Umsetzung gegeben! Der WWW-Ansatz akzeptiert, dass bestimmte Anforderungen nicht voll erfüllt werden können. Eine Annäherung an die Problemstellung ist in diesem Paradigma bereits eine Lösung. Dieser Ansatz ist logisch konsequent, da die Komplexitätsreduktion die vollständige Lösung nicht mehr gewährleistet.⁴² Im folgenden soll nun, mit der dargelegten Sichtweise als Basis, das Konzept und die Grenzen des „Semantic Web“ analysiert werden. Bei dieser Analyse ist zu berücksichtigen, dass die Paradigmen des WWW gelten, welche die alten „technikbasierten“ Paradigmen auf den Kopf stellen.⁴³

Die zentralen Punkte des WWW-Entwurfsparadigmas sind also:

- offene Standards

Spezifikation und Realisierung. Dabei kann das evolutionäre Entstehen eines Systems nur bedingt akzeptiert werden. Das Aufbrechen in kleine Teillösungen ist aber auch hier zu beobachten. Agile Entwicklungsmethoden wie „Scrum“ kommen der Idee der evolutionären Entwicklung sehr nahe. (Siehe Schiller/Canditt, 2008, S. 24-28), (auch Petrasch/Fieber/Bednarz, 2008, S. 36-44).

⁴¹Allgemein gesagt: Komplexitätsreduktion wurde notwendig, da die Summenkomplexität nicht beherrschbar war. Die Komplexitätsreduktion durch Teilsystembildung kann daher ihre Vollständigkeit nicht belegen. Die Summenkomplexität basierte auch auf der Unkenntnis von Eigenschaften (vgl. Luhmann, 1987, S. 50-51). Diese unbekannten Eigenschaften können daher nicht geplant erfasst sein. Auch handelt es sich bei großen Softwareaufgaben – gerade im Bereich der KI – um nichtdeterministische komplexe Systeme. Sie weisen daher eine Intransparenz (zur Problematik komplexer Systeme siehe auch Heyl, 2005, S. 132-145) auf. Die Komplexitätsreduktion erfolgt damit – ev. unbewusst – durch Reduktionismus.

⁴²Eigentlich: Ausschließt (siehe S. 17). Da zur Komplexitätserniedrigung das Ziel nur als grobe Vision gegeben ist, existiert keine erreichbare Zieldefinition. Der evolutionäre Entwicklungsprozess schließt aber ein funktionales Erreichen oder sogar eine Übererfüllung der ursprünglichen Zielvorstellung nicht aus. Evolutionäre Prozesse haben nichtlineare Charakteristika, die verlässliche Zukunftsprognosen verhindern und somit auch keine Entwicklung ausschließen können.

⁴³Die Frage, ob die Welt nun „Kopfsteht“ oder ob das WWW die EDV auf „die Füße gestellt hat“ soll nicht gestellt werden.

Gegenüberstellung der Paradigmen

Herkömmliches DV Konzept	Semantic Web Konzept
deterministisch	indeterministisch
funktional technisch	datenzentriert
aufgabenspezifisch	universell
umfassend spezifiziert	iterativ spezifiziert
institutionell realisiert	kollektiv, pluralistisch realisiert
prozedural gelöst	orchestriert
proprietär	offene Protokolle
	fehlertolerant

Tabelle 3.1: Vergleich alte / neue Paradigmen

Herkömmliches DB Konzept	Semantic Web Konzept
schemabasiert	verweisbasiert, dynamisch
redundanzfrei	redundant
hierarchische Relationen	Verweisnetze
konsistent	inkonsistent
persistent	flüchtig
semantische Integrität	widersprüchlich, mehrdeutig

Tabelle 3.2: Unterschiede Datenbank / WEB Anforderungen

- Einfachheit
- Fehlertoleranz
- Datensourcen kommunizieren mit Datenquellen
- Kommunikationsendpunkte sind Browser und Server
- möglichst keine Anforderungen an Datenquellen

Mit Web 3.0 wird diese paradigmatische Entwicklung fortgesetzt: Im Bereich der Interoperabilität will man von plattformunabhängig zu applikationsunabhängig gehen. Im Bereich des WWW will man vom Netz der Objekte (Dokumenten) zum Netz der Daten (Semantic Web) kommen.

4 Begriffsbildung und Bedeutung

4.1 Allgemeine Betrachtungen

Die Paradigmen des WWW haben auch Auswirkungen auf die Begriffs- und Benennungswelt im WWW. Dies gilt für angebotene Dienste und auch für die technologischen Konzepte.

Begriffe werden auch aus vermarktungstechnischen ⁴⁴ Gründen geschaffen - oder aus Profilierungsstreben ⁴⁵.

Es ist daher üblich, Leistungen oder Konzepte mit Schlagworten zu bezeichnen, die eher den Kriterien der Produktwerbung als den der Funktionsbeschreibungen genügen. Wurden zu Beginn noch Begriffe wie „Browser“ oder „Mail-Client“ verwendet, die noch entfernt die Funktion beschrieben,⁴⁶ so werden zunehmend Neologismen kreiert, die keinen unvorbereiteten Zugang ermöglichen.⁴⁷ Zwar ist es in Fachsprachen seit Langem üblich, Abkürzungen zu verwenden, die dem nicht Fachmann nicht ohne weiteres zugänglich sind. In der Welt des WWW sind Neo-

⁴⁴Vermarktungstechnisch ist hier nicht rein ökonomisch gemeint. Um seine Ideen im Web zu positionieren, muss man Strategien entwickeln, um bemerkt zu werden. Ein Beispiel für dieses Vorgehen ist „Lonelygirl15“ auf Youtube. Das Mädchen war weder 15 noch alleine. Die Videobeiträge waren auch keine Amateurvideos, sondern inszenierte Settings. Durch die Illusion, einem einsamen 15-jährigen Mädchen im privaten Metier zuzusehen, was durch die Benennung suggeriert wurde, erreichten die Videobeiträge Spitzenwerte in der Zugriffsstatistik.

⁴⁵Unter Profilierungsstreben ist hier das Erreichen von Einfluss auf die Entwicklungsrichtung gemeint. (Bsp. Kingsley, 2010) Dies ist auch im universitären Bereich wichtig, (siehe THESEUS, 2010) nicht zuletzt um zu Forschungsförderung zu kommen. Aber auch die kommerziellen Interessen - ev. im Bereich von Web-Services und Ontologien sind nicht auszuschließen. (siehe Fillies/Weichhardt, 2011; siehe OpenLink, 2010b). Außerdem gibt es noch zahlreiche Dienste und Firmen, die ev. nicht wirklich etwas mit dem Semantic Web zu tun haben, aber aus Gründen des Brandings Begriffe prägen. Kingsley selbst sagt in seinem Interview: „There are many waves to come, in a nutshell.“ (Kingsley, 2010)

⁴⁶„Browser“ von „brows“ = grasen, weiden, oder blättern (siehe Messinger/Rüdenberg, 1974, S. 88); Mail-Client = Post-Kunde

⁴⁷Bsp.: Googeln, Cookie, CiteULike, Twitter, Web 3.0, Wiki, Blog, Blogrole, LINES, FOX, Triplify usw.

logismen aber auch Teil des Lebensgefühls der „Community“ sowie zur Erzeugung von Markenzeichen im Gebrauch.⁴⁸ Ein funktionaler Zusammenhang zwischen Namen und Funktion ist nicht immer gegeben. So muss man eben wissen, dass mit „Google“ eine Suchmaschine gemeint ist, obwohl der Name eine Firma bezeichnet. Solche Gleichsetzungen hat es immer schon gegeben⁴⁹ - sie blieben aber eher auf die Umgangssprache oder die Produktwerbung beschränkt. Fachspezifisch wurde meist ein systematischerer Begriffsapparat verwendet.

Erschwert wird die Zugänglichkeit noch durch die Offenheit der Begriffe also die Vermeidung von Definitionen. Die Begriffe entstehen eben im Nutzungszusammenhang und dienen der Kommunikation innerhalb der Nutzergruppe. Dies ist Teil der Web-Idee. Dieses Phänomen der Ausbildung von anwenderspezifischen Sprachgewohnheiten, auch aus Gründen der Abgrenzung Aussenstehender gegenüber ist ebenfalls nicht neu. So war es unter anderem war auch beim Aufkommen der CB-Funk-Community in den 1970er Jahre zu beobachten.

Dieses Verhalten, ist eine immanente Eigenschaft von evolutionären, pluralistischen Prozessen. Es ist die Werbung für die gefundene Lösung. Etwa vergleichbar mit dem Rad, das der Pfau schlägt, um im Brautwerben auf sich aufmerksam zu machen. Pluralistische, anarchische Prozesse brauchen Mechanismen, um die Selektion der Lösungsvielfalt einzuleiten. Dazu müssen sie sich einerseits Abgrenzen (Gruppenbildung) andererseits auch auffällig werben.

Da das Gebiet von Semantic Web noch relativ jung ist, sind die Selektionsprozesse noch nicht abgeschlossen. Dadurch ist das Begriffs- und Lösungsfeld noch dementsprechend vielfältig. Lösungen und Begriffswelten sind eben oft noch eng mit Einzelgruppen verknüpft, die um Themendominanz ringen. Für Außenstehende ist das Fachvokabular daher nur sehr mittelbar mit Bedeutung verknüpft bzw. im Aussagegehalt schwer zuzuordnen⁵⁰. Die Begriffe benennen eben nicht ableit-

⁴⁸Soup.io, Teamdrive, Twitter, Yahoo usw. sind Beispiele.

⁴⁹Eine dieser Gleichsetzungen ist z.B. der „Dieselmotor“, bei dem der Begriffsinhalt nichts anderes sagt, als dass der Motor so ist, wie ihn Herr Diesel baut.

⁵⁰Als Beispiel aus einer Produktbeschreibung: „[...], SPARQL Create, Update, and Delete (SPARUL); [...]FOAF+SSL or OAuth based SPARQL endpoint security; Supports data broad range of RDF model data representation formats: HTML+RDFa, RDF-JSON, N3, Turtle, TriG, TriX, and RDF/XML; [...]“ (OpenLink, 2010a) Auch wenn alle die Begriffe Protokolle, Funktionalitäten usw. benennen, so ist doch die verwirrende Vielfalt und Begriffsbildung zu sehen. Eine ähnliche, nahezu Manie, alles mit kryptisierte Bezeichnungen zu versehen findet man sonst beim Militär (HEMTT für Lastwagen, HMRT für Anhänger, HMWWV für Geländefahrzeug usw.).

bare Eigenschaften wie zum Beispiel der Begriff „Automobil“, der etwas bezeichnet, was sich selbst bewegt. Oder „Dampflokomotive“, wo aus den Begriffsteilen die Funktionsprinzipien abgeleitet werden können - eben ein dampfbetriebenes Bewegungsmittel. Die Begriffe im WWW ähneln eher dem Aufbau von Marketingnamen wie etwa „VW Golf“ wo man keinesfalls schließen darf, dass hier ein „Wagen für das Volk für den Einsatz an einem Golf“ oder zum „Golf spielen“ gemeint ist. Es wird nur ausgedrückt, dass es im Themenbereich „VW“ irgendwas gibt, was „Golf“ heißt.

Wie vorsichtig man im Bereich des WWW mit Begriffen sein muss, sei hier anhand der OWL gezeigt. Das Acronym OWL hat nicht nur eine eigenartige Acronymbildung – „OWL“ statt „WOL“ – sondern wird im Namen als „Language“⁵¹ ausgewiesen, obwohl es sich hier meist um eine Architekturkomponente ohne Realisierungsfestlegungen handelt. „Meist“ bedeutet hier, dass OWL auch als realisierte Variante existiert und als solche oftmals auch einfach OWL genannt wird. Es ist eben beides: Konzept und auch Sprachdefinition (Bestandteil des Stack of Language). Die Begriffe verschwimmen hier ineinander. Systematischer wäre z.B. die Bezeichnung „OWA“⁵² für „Web-Ontology-Architecture“ und „OWL“ mit „L“ für Language als Protokollfestlegung für den Language Stack. Als Realisierung könnte dann ein Acronym wie etwa „XOWL“ verwendet werden.

Aber „könnte“ und „sollte“ ist hier nicht zielführend. Durch das auf Offenheit und demokratische Beteiligung zielende Paradigma der WWW-Community ist die Systematik zweitrangig. Es ist daher zu beachten, dass alle Abkürzung und Acronyme letztendlich inhaltsleer sind. Sie sind nichts anderes als Identifier für W3C-Spezifikationen.⁵³ Die Begriffswelt des WEBs ist keine Welt, die sich im philosophischen Sinn mit Begriffen für Objekte darstellen lässt. Es ist keine Welt,

⁵¹Sprache ist hier als Konstrukt aus definierter Syntax, Semantik und Vokabular gemeint – dies ganz im Sinne von Computersprachen. Auf eine andere Sichtweise zum Verständnis von „Language“ weist Sahle hin: „OWL steht in der Tradition der Auszeichnungssprachen. Auch diese nennt man Sprachen. Man nennt sie Sprachen, weil sie eine Terminologie und Regeln der Verwendung dieser Terminologie bieten. Insofern könnte man den ‚Sprach‘-Begriff auch unproblematisch finden.“(Patrik Sahle)

⁵²Noch konsequenter wäre natürlich WOA.

⁵³In diesem Sinn unterscheiden sich diese Begriffe nicht von der durch Referenz erzeugten Sinnggebung der Umgangssprache. Problematischer ist da schon, dass die Referenz selbst unklar ist. Die im Entstehungskontext (Context of discovery) gegebene Referenz wird häufig nicht geliefert, sodass erst im späteren Gebrauch Referenzen zu den Begriffen gebildet werden müssen. Darin unterscheiden sich viele dieser WWW Begriffe von sonst gebräuchlichen Begriffen. Sie werden publiziert, ohne klar anzugeben, was der intendiert Sinn ist. Die hier angegeben Referenz auf W3C Dokumente ist letztendlich eine von mir willkürlich gewählte Referenz.

in der vom Namen eines Objekts auf das Objekt geschlossen werden kann (Objektreferenz im philosophischen Sinn). Der Name (Identifier) verweist nur auf eine Beschreibung oder einen Beschreibungsteil (Aspekt) eines Objekts im WWW. Die humane Welt der WEB Community hat hier bereits vollzogen, was das „Semantic Web“ nun technisch nachvollzieht. Dies ist auch bei der technischen Realisierung des „Semantic Web“ zu berücksichtigen. Der Term „menschliche Sprache“ bezieht sich in der WWW Welt auf eine Realitätsdarstellung wie sie in der WWW Community gebräuchlich ist – nicht auf irgend eine Alltagsweltsprache. Das ist eben der Preis, den man für die Komplexitätsreduktion in der Modellbildung bezahlt: Die in einem Teil reduzierte Komplexität erhöht die Komplexität in anderen Teilen.

Was die Begriffe bedeuten, muss erst erschlossen werden. Diese Erschließen der Bedeutung kann auf zwei Wege erfolgen:

1. Erschließung aus bestehenden Dokumenten zum Thema.
2. Aktive Beschäftigung mit dem Thema und dadurch individuelle Bedeutungszuweisung.

Der zweite Weg, die Zuweisung eigener Bedeutung, ist bei neuen Themen angebracht. Je älter das Thema ist, um so mehr muss auf bereits bestehende Bedeutungen Rücksicht genommen werden. Es ist aber eine Eigenheit der pluralistische Web-Community, dass auch die Begriffswelt dynamisch bleibt und Umdeutungen oder Bedeutungserweiterungen erlaubt sind.

4.2 Der Begriff „Semantic Web“

Diese Begriffsproblematik betrifft auch die Wortschöpfung „Semantic Web“. Es ist nicht klar, was damit gemeint ist. „Ein Netzwerk von Bedeutungen“ - ja: aber was bedeutet das?

Versucht man in die Fachliteratur einzudringen, so ist zu sehen, dass auch dort die zugewiesenen Bedeutungen spartenspezifisch und interessenspezifisch sind. Ein Beispiel ist ein Artikel von Christian Fillies und Frauke Weichhardt, die im Semantic Web „die gemeinsame Nutzung mehr oder weniger formalisierten Wissens via *In-*

ternet.“⁵⁴ sehen. Die einen sehen die Aufgabe darin, Textinformation in ihrer Bedeutung maschinell zugänglich zu machen (Ansatz der künstlichen Intelligenz - KI), die anderen wollen die Interoperationalität verbessern.⁵⁵

Zur Klärung der Bedeutung von „Semantic Web“ versuche ich also vom Begriff selbst ausgehen.

[...] „Semantic Web“ kann auch als präziser und funktionierender Begriff verstanden werden. Das „Semantic Web“ ist ein „Web“, das sich vom normalen Web durch die Eigenschaft „Semantic“ unterscheidet. [...] In einem nächsten Schritt würde man nur noch untersuchen müssen, was mit diesem „semantic“ in Theorie und Praxis gemeint ist.⁵⁶

Dieser Analysezugang entspricht genau der Definition von Berners-Lee.⁵⁷ Er geht von den formalen Festlegungen aus und überlässt die Inhaltszuweisung des Begriffes „semantic“ der Begriffsdefinition. Das Problem ist aber, dass es diese Definition nicht gibt. Damit wird der Begriff offen.

Es ist hier die in der Einleitung⁵⁸ und auch später⁵⁹ dargelegte Problematik zu sehen. Es wird ein Begriff eingeführt, der aus Wörtern gebildet wird, die bereits Bedeutung tragen, um in einem zweiten Schritt diese Bedeutungen partiell aufzuheben. Der Bedeutungsinhalt von „Web“ bleibt erhalten, während „semantic“ für eine noch zu bestimmenden Funktionserweiterung als Platzhalter dient. Die durch „semantic“ verbundene Bedeutung als „[...] Bedeutungen sprachlicher Zeichen und Zeichenfolgen [...]“⁶⁰ soll nur metaphorisch angewendet werden.

Es ist eben so, wie im Zitat von Sahle am Kapitelanfang dargestellt: Der Begriff steht eben als Benennung ohne weiteren Bezug da. Auch wenn die Wortschöpfungen den Eindruck erwecken, sie hätten eine inhärente Bedeutung ist dem nicht so. Es soll zwar eine Denkrichtung vorgegeben werden, aber die genauere Bedeutung erschließt sich erst durch die Beschäftigung mit dem Thema. Zudem ist bereits seit den 1970er Jahren ein ähnlicher Begriff, nämlich der der „semantischen Netze“

⁵⁴Fillies/Weichhardt, 2011.

⁵⁵Vgl. Hitzler et al., 2008, S. 11.

⁵⁶Patric Sahle.

⁵⁷Vgl. Berners-Lee/Hendler/Lassila, 2001. Siehe auch Kap. 1 S. 4.

⁵⁸Siehe Kap. 1 S. 4.

⁵⁹Z.B.: Kap. 4 S. 19.

⁶⁰Dose, cop 1990, S. 709.

in Gebrauch.⁶¹

Für die Interpretationsmöglichkeit von Aussagen nachfolgend einige Beispiele. Als erstes die Ansichten über Semantic Web von Christian Fillies und Frauke Weichhardt, die sich mit der graphischen Erstellung von Ontologien beschäftigen:

Diese neue Technologie [Anm.d.Verf.: Semantic Web ist gemeint] wird sich zunächst innerhalb der Intranets größerer, verteilter Unternehmen ausbreiten, da hier ein großer Bedarf zur Abstimmung der Strukturen von Knowledge-Management-Systemen zwischen den verschiedenen Unternehmensteilen besteht, welche durch die Technologie des Semantic Web befriedigt werden kann. Die Schaffung zentraler Begriffskontexte bzw. Ontologien und Prozessmodellen ist dazu eine wesentliche Voraussetzung.

Die Idee des semantischen Webs ist die gemeinsame Nutzung mehr oder weniger formalisierten Wissens via Internet. Ursprünglich stand der Gedanke, eine gemeinsame Wissensbasis für Computer zu schaffen, im Vordergrund. Die hier beschriebene Arbeit behandelt jedoch zunächst den Wissensaustausch zwischen Menschen innerhalb eines (graphischen) Wissensnetzwerks.

Schlüsselfaktor für die verteilte Wissensnutzung ist die Verwendung einer gemeinsamen Terminologie.⁶²

Hier ist die Bandbreite der Interpretation zu sehen: Semantic Web als Intranet Interface für Knowledge-Management-Systemen⁶³ anstelle einer allgemeinen Ergänzung des WWW wie es Berners-Lee vermuten lässt⁶⁴, eine „gemeinsame Wissensbasis für Computer“⁶⁵ aber auch der Wissensaustausch zwischen Menschen mittels eines „(graphischen) Wissensnetzwerks“⁶⁶ als Thema von Semantic Web. Eine leicht andere Sichtweise ist bei Kingsley zu erkennen:

⁶¹Vgl. Fillies/Weichhardt, 2011. Inwieweit die Methoden und Begriffe der semantischen Netze auch auf das Semantic Web anzuwenden sind, wäre noch zu klären.

⁶²a. a. O.

⁶³a. a. O.

⁶⁴Vgl. Berners-Lee/Hendler/Lassila, 2001

⁶⁵Fillies/Weichhardt, 2011.

⁶⁶a. a. O.

A.B. [Anm. d. Verf.: Interviewer]: On some occasions you were pointing out the void between domain experts, their „internal“ ways to represent knowledge and the way this is done by traditional (database) technologies. Do you think that W3C’s specifications around the Semantic Web can offer a solution for this dilemma?

K.I. [Anm. d. Verf.; Kingsley Idehenn]: I don’t think this is an issue of specifications any more. The challenges lie in the alignment of terminology, and in generating an appreciation for the history of database technology. The envisioned Semantic Web is part of an innovation continuum that includes 40+ years of DBMS research.⁶⁷

Semantic Web wird als datenbankartige Datenansammlung betrachtet, für die es gilt, Zugriffsnormen und Kataloge zu erstellen. Kinsley ist CEO von OpenLink Software, die sich mit netzbasierten Datenbanken beschäftigt und so legt er seinen Betrachtungsschwerpunkt auf diese Facette.

Auch andere Autoren vertreten diese Sichtweise, dass Semantic Web im Kern aus Normierungen, Verzeichnissen und Strukturbeschreibungen (Dictionaries, Ontologien) besteht, welche die Interoperabilität zwischen Datenspeichern (Datenbanken, Wissensmanagementsystemen) ermöglichen sollen.⁶⁸

W3C – das Standardisierungsgremium für WWW schließt sich in ihrer Beschreibung von 2010 der Betrachtungsweise von Kingsley weitestgehend an:

In addition to the classic „Web of documents“ W3C is helping to build a technology stack to support a „Web of data,“ the sort of data you find in databases. The ultimate goal of the Web of data is to enable computers to do more useful work and to develop systems that can support trusted interactions over the network. The term „Semantic Web“^{*} refers to W3C’s vision of the Web of linked data. Semantic Web technologies enable people to create data stores on the Web, build vocabularies, and write rules for handling data. Linked data are empowered by technologies such as RDF, SPARQL, OWL, and SKOS.⁶⁹

⁶⁷Kingsley, 2010.

⁶⁸Vgl. Hitzler et al., 2008, S. 11.

⁶⁹W3C, 2010.

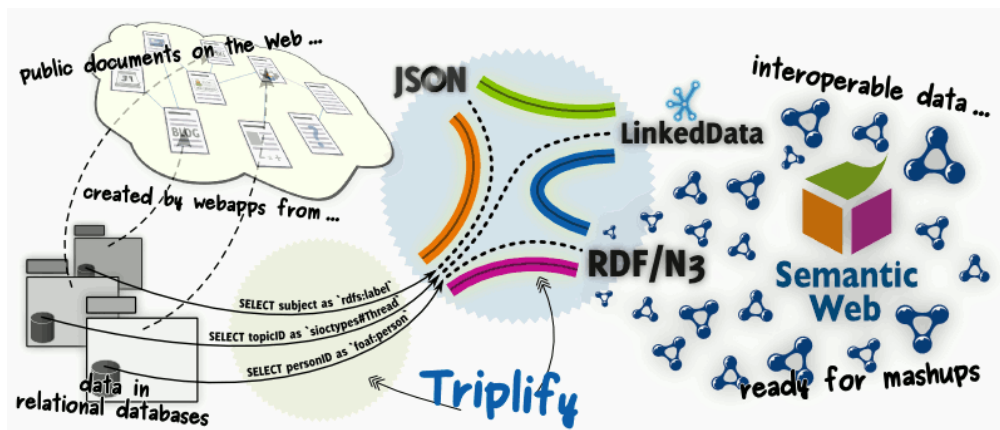


Abbildung 4.1: Semantic Web im Internet

(Quelle: Jaenicke, 2009)

W3C verwendet als Begriffe für das Web2.0 den Begriff „Web der Dokumente“ („Web of documents“⁷⁰) - also ein Netzwerk, in dem nach Dokumenten gesucht werden kann und für das Web3.0 den Begriff „Web der Daten“ („Web of data“⁷¹) gesehen als Datenbankinhalte. Inwieweit dies die ursprüngliche Vision verkörpert, ist unklar. Wie alles im Bereich WWW ist es schwer zu beurteilen, ob dann Semantic Web eine WWW Erweiterung ist, oder einer von vielen Diensten auf der Stufe von Suchmaschinen - oder ob das ohnehin ein und dasselbe ist.⁷² Auch die restlichen Erklärungen auf der W3C-Web-Page sind eher konventionell gehalten und haben kaum (mehr?) visionären Gehalt. Es scheint eben, dass W3C die Sichtweise von Kingsley übernommen hat. Selbst die maschinelle Nutzung, bei Berners-Lee noch ein zentraler Punkt, ist in der Darstellung des W3C in einen Klammereinschub gerutscht:

Using OWL (to build vocabularies, or „ontologies“) and SKOS (for designing knowledge organization systems) it is possible to enrich data with additional meaning, which allows more people (and more machines) to do more with the data.⁷³

Dass beim Versuch dem Begriff „Semantic Web“ Bedeutung zu geben, ohne ihn einzuschränken auch das eigentlich Gewollte, Spezifische, unbemerkt verloren

⁷⁰W3C, 2010.

⁷¹a. a. O..

⁷²Siehe Fig. 4.1 S. 26

⁷³W3C, 2010.

gehen kann, ist bei Hitzler zu sehen. Als Abschluss seiner Erklärung, was Semantic Web ist, schreibt er:

Die Ziele des Semantic Web sind eigentlich wesentlich moderater: Finde Wege und Methoden, Informationen so zu repräsentieren, dass Maschinen damit in einer Art und Weise umgehen können, die aus menschlicher Sicht nützlich und sinnvoll erscheint.⁷⁴

Diese Anforderung erfüllt bereits ein Sicherheitsventil.

Vetter weist auch darauf hin, dass es eben **keine** Definition für „Semantic Web“ gibt. Unter dem Begriff sollen alle Möglichkeiten verstanden werden, im WWW Inhalte maschinenverarbeitbar zu machen, also durch „semantischen Informationen (Metadaten), die Automatisierung [zu] ermöglichen.“⁷⁵

Die allgemeine Zusammenfassung der unterschiedlichen Aussagen ist: Alles was http-codierte Daten im WWW verarbeiten kann, ist Teil des „Semantic Web“. Damit sollte eigentlich für jede Sichtweise Platz sei - aber eben auf Kosten der Begriffsschärfe.

Allerdings ist die WWW-Community so weit anarchisch, dass sich jeder auch eine eigene Sichtweise bilden kann. Dies dann zu dem Preis, auch eine eigen Community aufbauen zu müssen um Aufmerksamkeit zu erlangen.

4.3 Was ist mit „Semantik“ gemeint

Wie vorher angerissen, ist in Semantic Web der Begriff „Semantik“ das unbestimmte Glied. Daher soll nun versucht werden, einen Zugang zu diesen Begriff zu finden. Dass die Begriffsklärung nicht so einfach ist, darauf weist eine mahnende Stellungnahme von Sahle hin:

Ich denke, dass man sich der Frage nach Semantik nicht nähern kann, ohne vorab festzustellen, dass sie Gegenstand verschiedener Fachrichtungen ist und dort unterschiedlich wahrgenommen wird. Und dass in

⁷⁴Hitzler et al., 2008, S. 12.

⁷⁵Vetter, 2006

dem Moment, in dem sich eine weitere Fachrichtung (wie die Informatik) sich dem Begriff nähert, sie wiederum eine eigene/andere Sicht entwickeln wird. Zu sagen, wie die Sprachwissenschaft Semantik versteht und dann festzustellen, dass dies bei der Informatik nicht ganz genau so funktioniert, ist doch nicht unbedingt zielführend.⁷⁶

Die große Vision von Semantic Web war, die im Web vorhandenen Daten inhaltlich in ihrer Bedeutung zu erfassen. Wie im Einleitungskapitel dargestellt, ist das Web auf eine textbasierte Kommunikation aufgebaut (http, HTML). Damit dürfte es der Grundidee (und dem Wort „semantic“) am nächsten kommen, zu ergründen, inwieweit Bedeutung aus dieser, für den Menschen gedachten Information gewonnen werden kann. Als Ausgangspunkt für Überlegungen scheint es mir daher sinnvoll, zu untersuchen, wie Bedeutung in der Sprachwissenschaft gesehen wird. Die Sichtweise der Technik oder noch schärfer, die der Informatik dürfte dagegen eher unbedeutend sein, da dort kaum andere Ansätze als die durch die Datenbanken⁷⁷ und domänenspezifische Kommunikationsprotokolle wie HL7⁷⁸ oder die verschiedenen SOA-Lösungen oder andere im Einsatz befindlichen Lösungen zu erwarten sind.⁷⁹

Semantik im Humanbereich	Semantik im technischen Bereich
mehrdeutig	eindeutig
offener Wahrheitswert	eindeutiger Wahrheitswert
Bedeutungsänderungen	konstante Bedeutung
dynamisches Vokabular	definierte Begriffe
kontextabhängig	Kontext als Typhierarchie
rückbezüglich	vorwärtsinterpretierbar
subjektive Interpretation	formal definiert, objektiviert

Tabelle 4.1: Semantikeigenschaften im Human- / Technikbereich

⁷⁶Patrik Sahle.

⁷⁷Gemeint sind hier die DB-üblichen Kriterien wie: Normalform, Eindeutigkeit, usw. Von der Benutzerseite vor allem die Forderung von semantischer Integrität. (vgl. Türker/Saake, 2006, S. 25)

⁷⁸HL7 = Health Level 7. HL7 ist ein Datenaustauschprotokoll im Spitalswesen. (siehe Wikipedia, 2011)

⁷⁹In der Softwaretechnik versucht man Semantik durch formale Definitionen festzulegen. „Die verschiedenen Forderungen nach *formaler Semantikbeschreibung* führten zu verschiedenen Ansätzen, wie operationale Semantik, funktionale Semantik, axiomatische Semantik und Übersetzungssemantik.“ (Jcobi, 1995, S. 335)

Ich gehe also davon aus, dass die Vision des Semantic Web darin bestand, möglichst direkt Bedeutungen aus den HTML und XML Daten im Netz zu erkennen.⁸⁰ Die Basis der Überlegung soll also der humansprachliche Semantikbegriff sein. Im *Handbuch wissenschaftstheoretischer Begriffe Band 3* ist Semantik wie folgt beschrieben:

S. ist die Wissenschaft der Bedeutung sprachlicher Ausdrücke.⁸¹

Semantik ist ein Teil der Semiotik.⁸² Semiotik befasst sich mit der Zeichenbe-
deutung an sich. Außerdem sind Syntax und Pragmatik Teile der Semiotik.⁸³ Für
den Anwendungsfall im EDV-Bereich scheint die Syntax als Aufbauregel wohlge-
formter Sätze⁸⁴ problemlos zu sein. Es wird hier daher nicht weiter darauf einge-
gangen. Die Pragmatik, also die Art des Sprachgebrauchs,⁸⁵ kommt erst in den
Fokus, wenn die Semantik geklärt ist. Daher soll vorerst einmal geklärt werden,
wie Semantik zu verstehen ist und die Pragmatik ausgeklammert werden.
Über die Aufgabe der Semantik ist bei Follesdal folgendes gesagt:

Die zwei Hauptfragen der S. sind „Wie bezieht sich die Sprache auf die
Welt?“ und „Welche Rolle spielt die Sprache in der Kommunikation?“⁸⁶

Um dies zu klären, soll vorerst der Begriff „Semantik“ näher analysiert werden.
In der Semantik sind zwei Begriffe von Bedeutung:

- Sinn
- Referenz

Die Referenz als der Verweis auf das Objekt scheint sowohl philosophisch als
auch DV-technisch weniger erklärungsbedürftig als die Sinnfrage.

⁸⁰Trotz dessen, dass die W3C - Seite zu Semantic Web nun einen anderen Eindruck vermittelt.(siehe W3C, 2010)

⁸¹Føllesdal, 1980, S. 568.

⁸²Die hier dargestellte Systematik richtet sich nach der Erklärung im *Handbuch wissenschaftstheoretischer Begriffe*.(Speck, 1980, S. 568) Es sei darauf hingewiesen, dass manche Theoretiker die Semantik anders einordnen.

⁸³Vgl. Føllesdal, 1980, S. 568

⁸⁴Vgl. a. a. O., S. 568.

⁸⁵Vgl. a. a. O., S. 568.

⁸⁶a. a. O., S. 568.

Auch die Beziehung zwischen einem Wort und seiner Referenz ist durch diese Kontroverse betroffen, da man von dieser Relation oft meint, sie basiert auf dem Sinn.⁸⁷

Aus diesem Blickpunkt soll die semantische Sinnfrage wie bei Føllesdal dargestellt wieder aufgenommen werden und zwar in der Variante, dass die Referenz den Sinn darstellt. Dies erfolgt aus dem Grund, dass Referenzen technisch einfach darzustellen sind und daher mit diesem Ansatz ein durchgängig plausibles Modell möglich wird.

Traditionelle Theorien der Referenz, wie z.B. die von Frege, erklären die Referenz mithilfe des Sinnes: die Referenz eines Ausdrucks ist jedes beliebige Objekt, auf den sein Sinn zutrifft.⁸⁸

Das könnte so verstanden werden, dass die Referenz auf ein Objekt der Sinn des Begriffes ist. Je vielfältiger diese Referenz ist, desto mehr Sinn ist erfasst.⁸⁹ Aber bereits das Herstellen einer einzigen Beziehung erzeugt Sinn.

Es ist hier zu beachten, dass in der EDV unter Referenz ein Verweis auf einen Speicherort zu verstehen ist während philosophisch der Bezug eines Objektbegriffs zu einem anderen Objektbegriff, letztendlich aber zum Sinnesobjekt steht. Der Speicherort kann als Objektinstanz, also als Analogie des Sinnesobjekts, verstanden und damit die Sichtweise jener der philosophischen Objekte gut angenähert werden.

Nun gibt es auch philosophische Darstellungen für Objekterklärungen, die ohne den Sinnbegriff auskommen,⁹⁰ doch scheint mir die auf Frege aufsetzende Erklärung des Sinns als Referenzidentität hier brauchbarer. Wenn die Referenz als Verweis auf ein Realweltobjekt verstanden wird, das EDV-mäßig durch einen Identifier dargestellt ist,⁹¹ so ist mit der Identifikation der Speicherreferenz (identer Verweis) ein Objekt identifiziert und somit erkannt.

⁸⁷Vgl. a. a. O., S. 569.

⁸⁸Vgl. a. a. O., S. 577.

⁸⁹Referenzen auf Kategorien, Klassenzugehörigkeiten usw. schafft immer mehr Sinn.

⁹⁰Siehe den Interpretationsversuch von Tarski, bei der jedem Ausdruck ein Wahrheitswert anstelle eines Sinnes zugeordnet wird. (vgl. Føllesdal, 1980, S. 577).

⁹¹Auch als Speicherreferenz zu interpretieren

Alternativ kann über zugeordnete Texteinträge (Objektbezeichnung) die Referenz gefunden werden. Es kann mit einer Objektbezeichnung – also einem Namen der auch als Anzeigetext dienen kann, Objektgleichheit hergestellt werden.⁹² Der Sinn könnte dann als erkannte Namensgleichheit⁹³ verstanden werden.⁹⁴

Im weiteren gehe ich aber zumindest von Tag-Namen in einem XML-Dokument als Objektbegriff aus.

Eintrag in Dokument 1:	<Objekt35> Baum <\Objekt35>
Eintrag in Dokument 2:	<ObjektBuche> Baum <ObjektBuche>

Daraus könnte als Sinn abgeleitet werden:

Objekt35 heißt Baum, ObjektBuche heißt Baum,
somit ist ObjektBuche und Objekt35 synonym.
somit könnte überall für Objekt35 vorkommt auch ObjektBuche eingesetzt werden.

Sinn-Erkennen in dieser Bedeutung würde sich auf die Abarbeitung eines Verzeichnisbaumes reduzieren, was DV-technisch leicht machbar ist.

Nun können im Prinzip zwei Referenzketten gebildet werden:

1. Horizontal: Alle Vorkommen des Eintrages auf der gleichen Ebene
2. Vertikal: Alle Vorkommen entlang von kategorischen Verwendungen (Ontologie)

Im ersten Fall würde ermittelt werden, welche unterschiedliche Namen das Objekt35 hat, im zweiten, welche Objekte als Baum bezeichnet werden. Da hier nur prinzipiell der Beleg für einfache Sinnerkennung durch Maschinen gezeigt werden sollte, möchte ich die Darstellung nicht weiter führen, zumal die aktuelle Realisierung durch RDF bereits weiter fortgeschritten ist.⁹⁵

⁹²Dies entspricht dem heutigen Suchen in Semantischen Netzen - oder in Google. Bei allen dieser Identifizierungsverfahren ist damit zu rechnen, dass immer Mehrdeutigkeiten, Mehrfachtreffer bestehen.

⁹³Gleichheit der Objektidentifikatoren, wobei den Objektidentifikatoren Anzeigetexte oder Bilder o.ä. zugeordnet sind. Namen sind eigentlich gleichwertig einer Speicherreferenz. Der Unterschied ist eher Realisierungstechnische: Namen in Freitexten unterliegen grammatikalischen Veränderungen.

⁹⁴Das Suchen und vergleichen von Bezeichnungen in freien Texten ist aber sehr problematisch. (vgl. Trunk, 2005, S. 12-15)

⁹⁵Siehe Hitzler et al., 2008, S. 81-85.

Auf zwei fehlende Eigenschaften dieser verkürzten Darstellung soll aber doch noch hingewiesen werden.

1. Art der Referenz

2. Die Klassifikation

Unter „Art der Referenz“ soll verstanden werden, was die gefundene Gleichheit aussagt. Im obigen einfachen Beispiel wurde dieser Referenzart nicht aus dem Text dekodiert, sondern willkürlich als „sein“ gesetzt. Damit wurde die sinnvolle Aussage: „Objekt35 ist Baum“ möglich. Es kann sich aber um die tatsächliche Aussage: „Objekt35 hat Baum“⁹⁶ handeln. Auch hier hat Semantic Web bereits einen weiterführenden Ansatz.

Unter „Klassifikation“ soll hier die Zuordnung zu Domänen und Kategorien verstanden sein. Umberto Eco unterscheidet zwischen Sinn und Bedeutung. Er sieht den Sinn in der „Art und Weise, auf die uns das Objekt gegeben ist.“⁹⁷ Die Referenz auf das Objekt sieht Eco dagegen als „Bedeutung“⁹⁸, also entgegengesetzt der Darstellung hier. Wichtig ist, dass das Erkennen der Zugehörigkeit zu Kontexten, zumindest zu Wissensdomänen letztendlich zur Sinnerkennung gehört. Dies allein schon aus Gründen der Begriffseindeutigkeit.⁹⁹

Im Zusammenhang mit dieser Art der Sinnerkennung, also das Erkennen welcher Kategorie ein Objekt angehört, wird im Semantic Web häufig der Begriff „Ontologie“¹⁰⁰ benutzt. Unter Ontologie wird dabei die Verweiskette der Referenzen verstanden. Auch für die Bearbeitung und Auswertung solcher Ontologien stellt die Semantic-Web-Technologie Festlegungen bereit.¹⁰¹

Die ursprüngliche Absicht, die per http transferierten Daten, mit wenig Metadaten angereichert, für Maschinen verstehbar zu machen, scheint möglich zu sein. Die Semantic-Web-Community und das W3C scheint aber zur Zeit den Weg zu gehen, doch ein umfassendes und vollständiges System spezifizieren zu wollen.¹⁰²

⁹⁶Beispiel: Email hat Nutzer.

⁹⁷Eco/Trabant, 1994, S. 70.

⁹⁸Vgl. a. a. O., S. 70.

⁹⁹Vgl. Trunk, 2005, S. 12-15.

¹⁰⁰„Lehre vom Sein, von den Ordnungs-, Begriffs- u. Wesensbestimmungen des Seienden.“(Dose, cop 1990, S. 551)

¹⁰¹Vgl. Herman, 2007; Vgl. DAML, 2010.

¹⁰²Vgl. W3C, 2010.

Es wird versucht, ein vollständiges Auszeichnungssystem mit Syntax und Semantik, etwa auf der Stufe eines serialisierten Datenbankschemas mit Typen, Klassen und Hierarchien zu erstellen.¹⁰³ Gleichzeitig werden aber weiterhin rudimentäre Ansätze unterstützt.¹⁰⁴

Der aktuelle Stand der Semantic-Web-Technologie hat damit sowohl für die erste Problemklasse¹⁰⁵ durch die RDF-Tripel¹⁰⁶, als auch für die Kategoriebildung¹⁰⁷ vor allem durch RDFS¹⁰⁸ und OWL¹⁰⁹ bereits ausgearbeitete Lösungen.

4.4 Schlussfolgerung

Semantik und somit Sinn im „Semantic Web“ kann prinzipiell direkt aus Web Dokumenten gewonnen werden. Sinn, basierend auf Objektidentifikation und auf der Abarbeitung von Verweisbäumen.¹¹⁰ Sinn wird dadurch erkannt, dass Objektgleichheit ermittelt wird. Wird Gleichheit festgestellt, können Aussagen wie „*Objekt1* ist ein *Basisobjekt*“ oder „*Objekt1* und *Objekt2* ist ein *Basisobjekt*“ getätigt werden. Abfragen wie: „Was ist eine Tanne?“ – „eine Tanne ist ein Baum“ werden damit möglich. Auf diese Art der Sinnermittlung lassen sich nun Logikverknüpfungen mit Wahrheitswerten für weitere Schlüsse und Urteile aufsetzen. Damit sollte diese einfache Interpretation des Sinnerkennens als atomare, einfache Basis für die maschinelle Verarbeitung im WWW ausreichen.¹¹¹

¹⁰³Vgl. Hitzler et al., 2008, S. 91-121.

¹⁰⁴Gemeint sind Ansätze, die möglichst einfache und wenige Auszeichnungselemente enthalten - etwa Mikroformate. (siehe HB, 2007)

¹⁰⁵Art der Identifikation.

¹⁰⁶Abbildung der Referenz durch jeweils drei Angaben: das Subjekt, das Prädikat, das Objekt. Das Prädikat gibt die Art der Referenz - also die Kanteneigenschaften des Bezugsgraphen an.

¹⁰⁷Zur Kategoriebildung existieren zumindest 2 Lösungen:

1. Namensräume (der Tag-Namen)
2. Verweisketten der Tripel

¹⁰⁸RDF-S = RDF Schema. Mit RDF-S lassen sich Hierarchien der Tripel modellieren.

¹⁰⁹OWL = Web Ontology Language. Die OWL dient zur Typisierung und Strukturierung der Einträge. In OWL wird die Hierarchie ausführlich modelliert (Erweiterte Möglichkeiten im Verhältnis zu RDF-S).

¹¹⁰Es sind noch weitere Sichtweisen denkbar – etwa: Sinnerkennen durch funktionale Abarbeitung etwa durch vorbereitete Fragmente, die durch Schlüsselbegriffe aktiviert werden. Dies Variante könnte aber als Spezialfall einer Referenz gesehen werden.

¹¹¹Im Semantic-Web-Toolset werden Interpretationen inzwischen auf formale Semantik - basierend auf der Satzlogik - aufgesetzt. (vgl. Hitzler et al., 2008, S. 91-121)

Abschließend können nun die beiden Eingangsfragen¹¹² zur Semantik allgemein wie folgt beantwortet werden:¹¹³

- Die Sprache des WEB bezieht sich über Identifikatoren auf die Welt
- Die Sprache dient der Verknüpfung von Identifikatoren in der Kommunikation.

Im Toolset des Semantic Web sind diese Semantikleistungen im RDF wie folgt gelöst:

- Die Identifikation durch URIs.
- Die Verknüpfung durch Prädikate (in den Tripel) und durch implizite Setzung.

Zusätzlich sind durch Namensräume domänenspezifische Unterscheidungen möglich. Eine Kategorisierung ist zum Beispiel durch RDF-S möglich.¹¹⁴

Die hier gegebene Interpretation von Sinn deckt sich auch mit anderen vorhandenen „Semantic Web“ Beschreibungen. So schreibt beispielsweise die SWAG¹¹⁵:

Das Semantic Web ist ein Netz, welches eine Menge von Dokumenten beinhaltet. Des Weiteren beschreibt es explizite Abhängigkeiten zwischen Teilen des Semantic Webs und beinhaltet semantische Informationen die für automatisierte Vorgänge von Maschinen vorgesehen sind. (Quelle: SWAG, Semantic Web Agreement Group)¹¹⁶

Die Durchgängigkeit der Interpretation von Semantic und Sinn sowohl in der Philosophie als auch in der technischen Umsetzungsmöglichkeit in Semantic Web sehe ich damit als gegeben an.

¹¹²Siehe Kap. 4.3 S. 29.

¹¹³Diese Antwort soll eine Semantikauffassung darstellen, die NICHT auf Technik fokussiert ist. Es soll eine Darstellung sein, die auch auf Umgangssprache zutrifft. Es soll in dieser Sichtweise noch die sogenannte *KI-Lösung* für die technische Abarbeitung nicht ausgeschlossen werden, auch wenn sich der Semantic-Web-Ansatz zum Ziel gesetzt hat, das Semantikproblem stark vereinfacht zu lösen.

¹¹⁴Gemeint ist hier das Property (Prädikatwert) aus der Datenmodellierung von RDF-S „subClassOf“. (siehe Semantic Web, 2007)

¹¹⁵SWAG = Semantic Web Agreement Group. (vgl. Schwartz, 2001).

¹¹⁶Zitiert nach Vetter, 2006.

5 Semantic Web Konzepte

Die Webinhalte in HTML oder XML bieten schon viele Möglichkeiten Inhalte Semantisch auszuwerten.¹¹⁷ So könnten etwa Name und Datum aus den vorhandenen Metatdaten des Kopf-Abschnitts eines XHTML-Dokuments, oder Mailadressen aus dem a-Tag erkannt werden.¹¹⁸ Andere Information könnten aus den Parameter des Div-Tag ¹¹⁹ gewonnen werden. Diese Einträge können nur mit großen Unsicherheiten genutzt werden. Um eine definiertes Erkennen von Bedeutungen zu ermöglichen, wurde in Semantic Web der Weg gewählt, eine eigene Semanticonotation, die in http-codierten Dokumenten verwendet werden sollen (Bsp. in HTML als Attributnamen oder in XML als Tags, aber auch in eigene RDF-oder OWL-Dokumenten.) zur Semantikauswertung festzulegen.

Eine eigene Syntax zur Objektidentifizierung und Ontologiedarstellung (Semantische Struktur der Daten) wurde spezifiziert. Dies ist eine weitere Komplexitätsreduktion des semantischen Analyseproblems. Dieses Konzept basiert nicht darauf, beliebige WEB-Dokumente in ihrer Semantik zu analysiert, sondern es wird eine klare Syntax und Semantik, die maschinenverarbeitbar ist – in Analogie zu Programmier- oder Datenbankabfragesprachen – geschaffen und vorhandene Dokumente damit instrumentiert oder Dokumente nach den neuen Richtlinien neu erstellt.¹²⁰ Inwieweit dies eine Einschränkung im Verhältnis zu Berners-Lee Ursprungsvision ist, ist offen.¹²¹ Es wird also ein Rahmenwerk für die syntaktische

¹¹⁷Siehe Basisüberlegungen Kap. 4.3 S. 27.

¹¹⁸Siehe Code-Beispiel Zeile 5-13 5.3 S. 46.

¹¹⁹Siehe Code-Beispiel Zeile 15 und 18 5.3 S. 46. Im Beispiel ist im Div-Tag ein Identifizierungsname vergeben. Allerdings kann ein Auswerteprogramm nicht erkennen, wofür die Identifikation steht.

¹²⁰Siehe dazu die Überlegungen in Kap. 4.3 S. 27.

¹²¹Aus meiner Sicht, wird mit diesem Lösungsansatz das innovative Potential der Vision untergraben. Der Lösungsweg, wie er sich aktuell in W3C darstellt, ist dem Paradigmenwechsel des WWW nicht adäquat. Ein Framework zur Handhabung serialisierter Datenbanken ist zwar kommerziell interessant - innovativ ist es aber nicht unbedingt.

Verarbeitung festgelegt.¹²² Diese Festlegung betrifft eine konzeptionelle und eine konkrete Komponente. Die Konzeption sind die Architekturmodelle, die Modelle der Objektdarstellung und die Semantikmodelle. Die konkreten Komponenten umfassen die Sprachdefinitionen in Syntax und Semantik und die Realisierungen.¹²³

5.1 Die Basiskonzepte

Auf der W3C-Homepage werden zu Semantic Web folgende Semantic-Web-Komponenten genannt:¹²⁴

- **Linked Data:** „The Semantic Web is a Web of data – of dates and titles and part numbers and chemical properties and any other data one might conceive of. RDF provides the foundation for publishing and linking your data. Various technologies allow you to embed data in documents (RDFa, GRDDL) or expose what you have in SQL databases, or make it available as RDF files.“¹²⁵
- **Vocabularies:** „At times it may be important or valuable to organize data. Using OWL (to build vocabularies, or ‚ontologies‘) and SKOS (for designing knowledge organization systems) it is possible to enrich data with additional meaning, which allows more people (and more machines) to do more with the data.“¹²⁶
- **Query:** „Query languages go hand-in-hand with databases. If the Semantic Web is viewed as a global database, then it is easy to understand why one would need a query language for that data. SPARQL is the query language for the Semantic Web.“¹²⁷
- **Inference:** „Near the top of the Semantic Web stack one finds inference – reasoning over data through rules. W3C work on rules, primarily through RIF

¹²²Siehe W3C, 2004.

¹²³Praxisbeispiele sind am Web unter <http://semanticdreamer.com/semantic-web-schnell-und-kompakt/> als Ergänzung zu dem Buch *Semantic Web* (Geisler, 2009).

¹²⁴Siehe W3C, 2010.

¹²⁵a. a. O..

¹²⁶a. a. O..

¹²⁷a. a. O..

and OWL, is focused on translating between rule languages and exchanging rules among different systems.“¹²⁸

- Vertical Application: „W3C is working with different industries – for example in Health Care and Life Sciences, eGovernment, and Energy – to improve collaboration, research and development, and innovation adoption through Semantic Web technology. For instance, by aiding decision-making in clinical research, Semantic Web technologies will bridge many forms of biological and medical information across institutions.“¹²⁹

5.1.1 Die Architektur

In den Beschreibungen ist manchmal der Übergang vom Abstrakten zum Konkreten nicht klar erkennbar. Dieser Umstand geht darauf zurück, dass die Systemarchitektur von Berners-Lee von Anfang an vier Referenzarchitekturen aufwies und damit für Verwirrung sorgte.¹³⁰ Zudem wurden diese Architekturmodelle nie als Teil der W3C Recommendation publiziert und so in ihrer Bedeutung konkretisiert.¹³¹

Gerber versuchte eine „Semantic Web“ - Systemarchitektur zu erstellen, die auf Softwareengineering-Prinzipien fußt. Er nannte sie CFL-Architektur¹³²

Gerber versucht nun im Vergleich mit seiner CFL die Architekturmodelle von Berners-Lee bzw. der W3C zu analysieren. Als Beispiel ist hier die Analyse des V4 Modells von Berners-Lee gezeigt.¹³³

In dem Vergleich des CFL-Modells mit dem Architekturentwurf Berners-Lee taucht auch wieder die konzeptuelle Verschwommenheit auf. Die Komponente „SPARQL“¹³⁴ ist nicht Bestandteil dessen, was Semantic Web ausmacht. „SPARQL“ ist für Semantic Web im Kern nicht nötig und daher gestrichen. Die Funktionalität an sich ist im Architekturmodell bereits in den Komponenten „RDF“,

¹²⁸a. a. O.

¹²⁹a. a. O.

¹³⁰Siehe Fig. 5.1 S. 38.

¹³¹Vgl. Gerber/van der Merwe/Barnard, 2007, S. 2-7.

¹³²CFL = Comprehensive, Functional, Layered; (Gerber/van der Merwe/Barnard, 2007, S. 6). (siehe Fig. 5.2 S. 39).

¹³³Siehe Fig. 5.3 S. 40.

¹³⁴Siehe W3C, 2007.

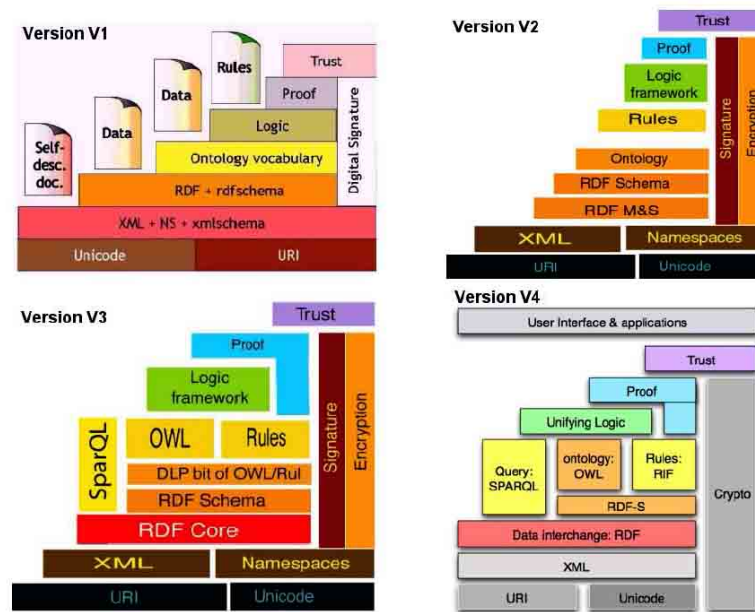


Fig. 1. The four versions of the Semantic Web reference architecture (V1-V4) proposed by Berners-Lee [6, 9–13].

Abbildung 5.1: Die 4 Versionen der Referenzarchitektur

(Quelle: Gerber/van der Merwe/Barnard, 2007, S. 3)

„Ontology“, „Rules“ und „Logic Framework“ enthalten.

Bei „SPARQL“ handelt es sich um das Konzept und die Realisierung einer Abfragsprache.¹³⁵ Sie ist Bestandteil der Nutzung von Semantic Web.

Hitzler fasst die Modellierung des „Semantic Web“ so zusammen:

[Es] werden Daten mit einer Struktur (Syntax) und Bedeutung (Semantik) versehen. In diesem Zusammenhang spricht man davon, Daten maschinenlesbar und -interpretierbar bereitzustellen. Eine herausragende Rolle spielen dabei die W3C1-Standards [...] XML, RDF(S) und OWL. Ein wichtiger Grundgedanke dabei ist es, das „Semantic Web“ als Erweiterung des bestehenden Web zu etablieren. Daher werden „Semantic Web“ – Standards auf bestehende Web-Technologien, wie z.B. URIs, aufgesetzt. Die einfache und schrittweise Erweiterung

¹³⁵Ich sehe darin eine Abfragesprache für „serialisierte Datenbanken“. Für ein Summenprodukt „Semantic Web“, das sich eben als „Datennetzwerk“ sieht, das auch vermarktet sein soll, ist SPARQL ein zwingender Systembestandteil.

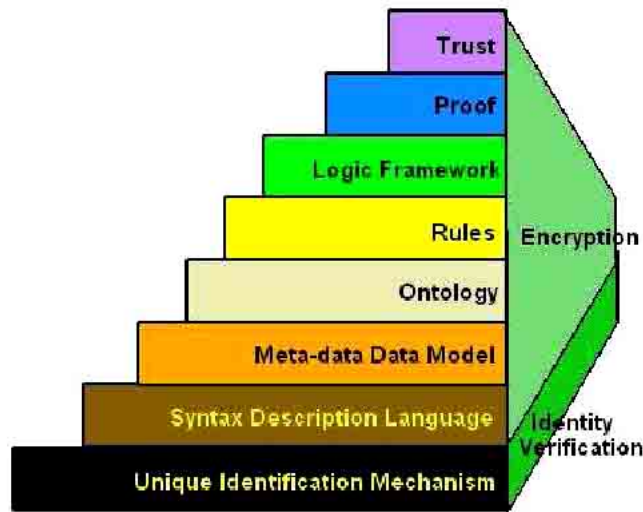


Fig. 3. The proposed CFL architecture for the Semantic Web comprises two orthogonal architecture stacks, the *language* stack and the *security* stack.

Abbildung 5.2: CFL-Architekturmodell von Gerber

(Quelle: Gerber/van der Merwe/Barnard, 2007, S. 7)

von bestehenden und auf bewährten Web-Technologien basierenden Anwendungen mit neuen „Semantic Web“ – Funktionalitäten soll so weit wie möglich gewährleistet bleiben.¹³⁶

Es ist in dieser Sicht für mich zu sehen, dass der interpretative Zugang zu vorhandenen Datenbeständen verlassen wird.¹³⁷

Die für das „Semantic Web“ notwendige WWW Ergänzung ist die „Ontologie“¹³⁸ genannte Ergänzung und eine Abfragesprache. WWW Ergänzungen im Überblick:¹³⁹

¹³⁶Einleitung Hitzler et al., 2008, S. 1.

¹³⁷„Daten werden mit [...] Struktur [...] und Bedeutung [...] versehen.“ (Einleitung a. a. O., S. 1). Dies widerspricht nach meiner Auffassung Berners-Lees Vision von einer WWW-Erweiterung. Die Nutzung des Semantic Web läuft mit diesem Ansatz entweder auf einen Ersatz des jetzigen WWW hinaus oder Semantic Web ist auf der Stufe eines Web-Service, wie sie auch heute schon durch RSS, Cloud Computing oder Google verwirklicht sind. Da, wenn Semantic Web den eingeschlagenen Weg weiter geht, dies eine neue Sprachdefinition und die Neuerstellung von Inhalten bedingt, können bestehende Daten nicht genutzt werden. Oder eben nur durch eine sogenannte „Migration“, wie es in der klassischen Datenbanktechnik üblich ist. Die Vision, das Problem der KI auf pluralistische Weise iterativ zu bewältigen, beginnt sich jedenfalls im Pool der pragmatischen Lösungen aufzulösen.

¹³⁸Einleitung a. a. O., S. w.

¹³⁹Die W3C Sichtweise ist die Basis dieses Überblicks.

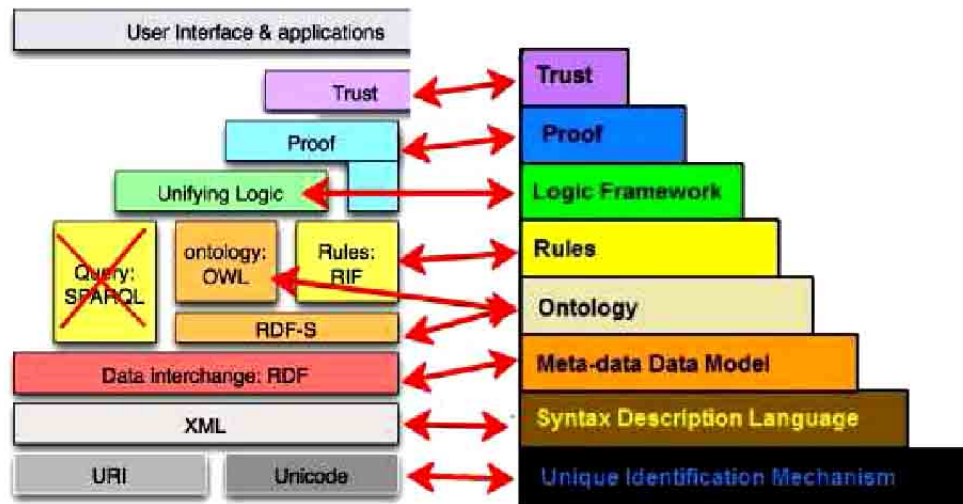


Fig. 6. The V4 version of the Semantic Web reference architecture [13] as instantiation of the adapted architecture.

Abbildung 5.3: Zuordnung Berners-Lee V4 Modell zu CFL-Modell

(Quelle: Gerber/van der Merwe/Barnard, 2007, S. 10)

- Ontologie (Erweiterte RDF¹⁴⁰, als OWL¹⁴¹ bezeichnet)
- Abfragesprache (RDF; Adaptierter SQL-Dialekt¹⁴² SPARQL¹⁴³)

Eine Übersicht über die „Semenatic Web“ – Architektur findet man bei Gerber et. al.¹⁴⁴

5.1.2 Die semantischen Konzepte

An semantischen Anreicherungskonzepten stehen im Grunde drei Ebenen zur Verfügung:

1. Nettodaten mit Attributen

¹⁴⁰RDF = Resource Description Framework.

¹⁴¹OWL = Web Ontology Language

¹⁴²SQL = Structured Query Language.

¹⁴³SPARQL = SPARQL Protocol and RDF Query Language. Streng genommen sollt SPARQL nicht notwendig sein. Das Abarbeiten und Auswerten sollte auch durch Verfolgen der RDF Tripel möglich sein.

¹⁴⁴Gerber/van der Merwe/Barnard, 2007, Siehe

2. Instrumentierung mit Objektreferenzen

3. Instrumentierung mit Semantikstrukturinformation

Die Methode der Anreicherung von Nettodaten mit Attributen ist besonders bei HTML geeignet. Die Instrumentierung mit Objektreferenzen ist das Kernstück von RDF. Die Instrumentierung mit Semantikstrukturinformation kann mit RDF-S oder OWL erfolgen.

Zur Festlegung eines Vokabulars werden zwei Wege beschritten:

1. Festlegung per „Norm“¹⁴⁵
2. Festlegung per Auflistung in zentralen Katalogen¹⁴⁶.

Im Semantic Web ist die Information ein sich selbst beschreibendes Netzwerk, das per Verweise – URIs – abgearbeitet werden kann. Es wird lediglich die Syntax eines Eintrages definiert. Jeder Eintrag ist mittels eines Tripels aus Subjekt-Attribut-Objekt beschrieben. Eine Hierarchisierung ist auch ohne OWL Modellierung durch die Objektverweise vorhanden. Es gäbe die Möglichkeit, sie prozesshaft abzuarbeiten und durch programmierte Algorithmen zu rekonstruieren.

Mit OWL besteht die Möglichkeit der Strukturfestlegung und der Objektkardinalität - also eine an die DB-Welt angenäherte Beschreibung. Die Beschreibung der Datensemantik durch die Ontologie-Einträge ähnelt zwar einem Datenbankschema und ermöglicht auch datenbankähnliche Abfragen, ist aber kein Schema im Sinne eines Datenbankschemas.¹⁴⁷

Damit ist es möglich (und auch beabsichtigt) Ontologien zu schaffen, welche serialisierten Datenbanken entsprechen - allerdings: Verpflichtend ist das nicht. Wird die Ontologie nicht in dieser Weise aufgebaut, können eben nicht alle Möglichkeiten - etwa SPARQL - genutzt werden. Gültige Datenbestände lassen sich trotzdem aufbauen.¹⁴⁸

¹⁴⁵Norm ist hier als Vereinbarung zu verstehen. W3C ist kein staatliches Normungsinstitut.

¹⁴⁶Etwa Dublin Core(Dublin Core Metadata Initiative, 2010).

¹⁴⁷Mit OWL kann eine Datenstruktur erzeugt werden, die einem serialisierten Datenbankschema entspricht. Der Unterschied ist aber, dass keine vollständige OWL Datenauszeichnung vorliegen muss, um überhaupt Daten speichern zu können. Datenbeschreibung und Dateninhalt greifen bei Semantic Web ineinander.

¹⁴⁸Datenbanken ohne Schemadefinition können nicht in Betrieb genommen werden. In WWW ist aber ein HTML konformes Dokument ein gültiges Dokument - unabhängig von der Ontologie-Definition. Das meine ich mit der Aussage, dass Semantic Web kein Schema benötigt.

5.2 Die Notationen

Die Umsetzungen der Konzept in konkreten Notationen ist nicht abgeschlossen. Es entstehen laufend neue Entwürfe und Realisationen.

Im allgemeinen wird bei den Notationen folgendes Konzept verfolgt:

- Eine Basisvariante
- Spezifische Varianten (mächtiger, oder vereinfacht)
- Notationen für andere Sprachen

Die Basisvariante ist meist in einer eigenen Syntax und deckt die unbedingt notwendigen Leistungsumfang ab.¹⁴⁹

Die spezifischen Varianten sind Spracherweiterungen, Notationsvereinfachungen o.ä.¹⁵⁰

Notationen für andere Sprachen sind Festlegungen, den Sprachumfang in anderen Protokollen zu nutzen.¹⁵¹

RDF beschreibt dabei drei Bereiche:

1. Ein Vokabular (RDF-Bezeichner).
2. Eine Syntax für RDF-Einträge (Bsp. RDF/XML: Aufbau der Einträge).
3. Eine Definition für Objektbezeichnungen (URI als Objektidentifikation).

Die RDF-Tripel können XML konform dargestellt werden¹⁵² – daher auch die Benennung RDF/XML.¹⁵³

In ihrer semantischen Bedeutung gibt es dabei mit bereits vorhandenen XML-Einträgen Überschneidungen - Bsp. bei den Namens-Tag. Allerdings ist der RDF-Tag präziser in seiner Verwendung als der relativ frei zu verwendende Eintrag von

¹⁴⁹Bsp.: RDF, einfache Mikroformate.

¹⁵⁰Bsp.: Zusammengesetzte Mikroformate, Turtlesyntax, RDF-S usw.

¹⁵¹Bsp.: RDF/XML, Verwendung von Dublin Core in HTML oder XML usw. Als Bsp. siehe Listing 5.1 S. 43(Geisler, 2009, S. 47-49) und Listing 5.2 S. 44(a. a. O., S. 65-67).

¹⁵²Ein Beispiel einer RDF/XML Notation ist in Listing 5.4 S. 47 gezeigt. Das Beispiel ist der W3C-RDF-Primer-Dokument entnommen.(siehe W3C, 2004, S. 4) Der RDF Graf wird dabei serialisiert.

¹⁵³RDF/XML ist aber nur eine Variante, semantische Information in Web-Dokumenten zu notieren. Andere sind: Tripel-Syntax, Notation 3, N-Triples, Turtle.(siehe Geisler, 2009, S. 27-29)

```

1 <!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.1//EN"
2 "http://www.w3.org/TR/xhtml11/DTD/xhtml11.dtd">
3 <html xmlns="http://www.w3.org/1999/xhtml" xml:lang="de">
4   <head>
5     <meta http-equiv="Content-type" content="text/html; charset=utf-8" />
6     <title>Semantic Web: schnell+kompakt</title>
7   </head>
8   <body>
9     <!--h1>Zusammengesetztes Mikroformat (h Card)</h1-->
10    <div class="vcard">
11      <a class="fn organization-unit url" href="http://entwickler-press.de/">entwickler.press</a>
12      <div class="organization-name">Software & Support Verlag GmbH</div>
13      <div class="adr">
14        <div class="street-address">Geleitsstraße 14</div>
15        <span class="country-name">DE</span>
16        <span class="postal-code">60599</span>
17        <span class="locality">Frankfurt</span>
18      </div>
19      <div>Tel:
20        <span class="tel">+49 (0)69 - 63 00 89 0</span>
21      </div>
22      <div class="tel">
23        <span class="type">Fax</span>:<span class="value">+49 (0)69 - 63 00 89 89</span>
24      </div>
25      <div>E-Mail:
26        <span class="email">info@entwickler-press.de</span>
27      </div>
28    </div>
29  </body>
30 </html>
31
32
33 Quelle: Geisler. 2009

```

Listing 5.1: Beispiel einer XHTML-Datei mit h Card Mikroformat

```

1
2 <!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.1//EN"
3   "http://www.w3.org/TR/xhtml11/DTD/xhtml11.dtd">
4 <html xmlns="http://www.w3.org/1999/xhtml" xml:lang="de">
5   <head profile="http://dublincore.org/documents/dcq-html/">
6     <meta http-equiv="Content-Type" content="text/html; charset=utf-8" />
7     <title>Semantic Web schnell+kompakt</title>
8     <!-- Expressing Dublin Core in HTML/XHTML meta and link elements ,
9           http://dublincore.org/documents/dcq-html/ -->
10    <link rel="schema.DC" href="http://purl.org/dc/elements/1.1/" />
11    <link rel="schema.DCTERMS" href="http://purl.org/dc/terms/" />
12    <meta name="DC.contributor"
13          content="Dr._Christoph_Schmidt" />
14
15    <meta name="DC.coverage"
16          content="Germany"
17          scheme="DCTERMS.TGN" />
18    <meta name="DC.creator"
19          content="Matthias_Geisler" />
20
21    <!-- [ Beispiel verkürzt ] -->
22
23    <meta name="DC.rights"
24          content="http://creativecommons.org/licenses/by-nc-sa/3.0/de/" />
25    <meta name="DC.source"
26          content="http://dublincore.org/documents/dcq-html/" />
27    <meta name="DC.subject"
28          content="Beispiel_zur_Anwendungen_von_Dublin_Core_Metadaten-Elementen." />
29    <meta name="DC.title"
30          content="Semantic_Web_schnell+kompakt" />
31    <meta name="DC.type"
32          content="Text"
33          scheme="DCTERMS.DCMIType" />
34  </head>
35  <body></body>
36 </html>
37
38 Quelle: Geisler.2009

```

Listing 5.2: Beispiel einer HTML-Datei mit DC-Einträgen

XML.

Der Semantic-Web-Ansatz setzt also auf eine Instrumentalisierung der Information mit Metadaten um die sichere Interpretation zu erleichtern. Dies ist eine im WWW gängige Technik. Auf ähnliche Weise wird auch versucht, Literatur in Bibliotheken maschinell zugänglich zu machen.¹⁵⁴

Mit diesen Schritten wird die Identifikation und Strukturierung der Daten möglich.¹⁵⁵

W3C schreibt dazu:

XML is not an easy tool for data integration. On the other hand, RDF consists of a very loose set of relations (triples). Due to its usage of URIs it is very easy to seamlessly merge triple sets, ie, data described in RDF within the same application; it is therefore ideal for the integration of possibly heterogenous information on the Web. But this has its price: reconstructing hierarchies from RDF may become quite complex. As an example, it would be fairly complicated (and unnecessary) to describe, eg, vector graphics, using RDF; use SVG instead!¹⁵⁶

An zusätzlicher Funktionalität werden benötigt:

- Eine Syntax für die Semantikeinträge.
- Eine Semantikmodellierung - Ontologie genannt.

In W3C ist die Basis der Semantikmodellierung die Spracherweiterung von RDF zu RDF-S (RDF Schema). Mächtiger ist die Ontologie-Modellierungssprache OWL.¹⁵⁷

Im Semantic-Web des W3C setzen die Ontologie - Funktionen auf das Konzept der RDF auf.

¹⁵⁴Es sei hier auf die TEI - Initiaive als Beispiel verwiesen.(Siehe TEI, 2011)

¹⁵⁵Zwar gibt es auch in XHTML und XML die Möglichkeit, Datenbeschreibungen durch Tags durchzuführen, allerdings sind diese Tagnamen nicht normiert. HTML, XHTML und XML bieten auch die Möglichkeit Objektverweise über Links in Datenströme einzubauen. In RDF ist diese Objektbeschreibung als Tripel systematisch geordnet.

¹⁵⁶Herman, 2009.

¹⁵⁷Bsp. eines OWL-Fragments siehe Listing 5.5 S. 47.

```

1 <?xml version="1.0" encoding="UTF-8"?>
2 <!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
3     "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
4 <html xmlns="http://www.w3.org/1999/xhtml" lang="de" xml:lang="de">
5   <head>
6     <link rel="stylesheet" type="text/css" href="./css/baseform.css" media="screen"/>
7     <meta http-equiv="Content-Type" content="text/html; charset=utf-8"/>
8     <meta name="author" content="W. Friedhuber"/>
9     <meta name="date" content="21.6.2010"/>
10    <meta name="description" content="Abschlussarbeit in 521.023 SS210 KFU"/>
11    <meta name="description" content="Erstellen einer Index.html mit Nutzung CSS"/>
12    <title>Abschlussprojekt 521.023</title>
13  </head>
14  <body>
15    <!-- Bereich der 'Nutzinformation' -->
16    <div id="main">
17      <!-- Kopfbereich: Überschrift und Photo -->
18      <div id="header">
19        <h1>Leitseite: Wolfgang Friedhuber</h1>
20        <p class="headtextblock">
21          <br/>
22          
23          Student an der Karl Franzens – Universität in Graz<br/>T
24          eilnehmer Kurs 521.023: "Webdesign" im SS2010 <br/>
25          <a href="mailto:w.friedhuber@edu.uni-graz.at"> Mailkontakt</a><br/>
26          <a href="http://www.stud.uni-graz.at/~74Friedh/"> Internetseite</a>
27        </p>
28      </div>
29    </div>
30  </body>
31

```

Listing 5.3: Beispiel einer XML-Datei

RDF based vocabularies, and the accompanying semantic formalisms like RDFS or OWL, also make it easy to define inference possibilities on RDF data. Although this could be done around XML dialects, too, it would remain application specific and not portable.¹⁵⁸

Bei all diesen Begriffen handelt es sich zuerst um Basiskonzepte, die noch nicht auf Realisierungsebene gedacht werden sollten. Es wird zuerst eine plattform- und realisierungsunabhängige Vorstellung entwickelt, die dann durch unterschiedliche Realisierungen konkretisiert werden.¹⁵⁹ Leider ist dadurch manchmal nicht mehr genau zu erkennen, wann vom abstrakten Modell und wann von der konkreten Realisierung gesprochen wird. So muss z.B. für das abstrakte Modell noch kein definiertes Begriffslexikon existieren – andererseits ist für Teile der Konzepte ein Vokabular und eine entsprechende Spezifikation für Metadaten (also eine Syntax) der Kern der Interoperabilität.

¹⁵⁸Herman, 2009.

¹⁵⁹Dieses Vorgehen wurde und wird auch bei anderen komplexen Problemen gewählt. So wurde CORBA (siehe Lena, 2011) ähnlich entwickelt.

```

1
2 <?xml version="1.0"?>
3 <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
4     xmlns:contact="http://www.w3.org/2000/10/swap/pim/contact#">
5
6   <contact:Person rdf:about="http://www.w3.org/People/EM/contact#me">
7     <contact:fullName>Eric Miller</contact:fullName>
8     <contact:mailbox rdf:resource="mailto:em@w3.org"/>
9     <contact:personalTitle>Dr.</contact:personalTitle>
10   </contact:Person>
11 </rdf:RDF>

```

Listing 5.4: Beispiel RDF/XML-Datei

```

1
2 <?xml version="1.0" encoding="UTF-8"?>
3 <!-- Quelle: gekürzt von http://www.w3.org/TR/owl-guide/wine.rdf -->
4 <rdf:RDF
5   xmlns="http://www.w3.org/TR/2003/PR-owl-guide-20031209/wine#"
6   xml:base="http://www.w3.org/TR/2003/PR-owl-guide-20031209/wine#"
7   xmlns:owl="http://www.w3.org/2002/07/owl#"
8   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
9   xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
10  xmlns:ex="http://example.com/">
11   <owl:Ontology rdf:about="">
12     <rdfs:comment>Eine Beispiel OWL-Ontologie</rdfs:comment>
13     <owl:versionInfo>v0.1</owl:versionInfo>
14     <owl:priorVersion>
15       <owl:Ontology rdf:about="http://example.com/example1"/>
16     </owl:priorVersion>
17     <owl:imports rdf:resource="http://example.com/example2"/>
18     <rdfs:comment>Ein beliebiger Kommentar...</rdfs:comment>
19     <rdfs:label>Beispiel Ontologie</rdfs:label>
20   </owl:Ontology>
21 </rdf:RDF>
22
23 Quelle: Geisler.2009

```

Listing 5.5: Beispiel einer OWL Datei

6 Die Technik

Die Basis des „Semenatic Web“ ist das bestehende WWW mit folgenden Eigenschaften (siehe Fig. 6.1 S. 48):

- Verwendung der Transportmechanismen des WWW (Protokoll, Adressierung)
- Anwenderinteraktion (Presentation Layer) über Browser

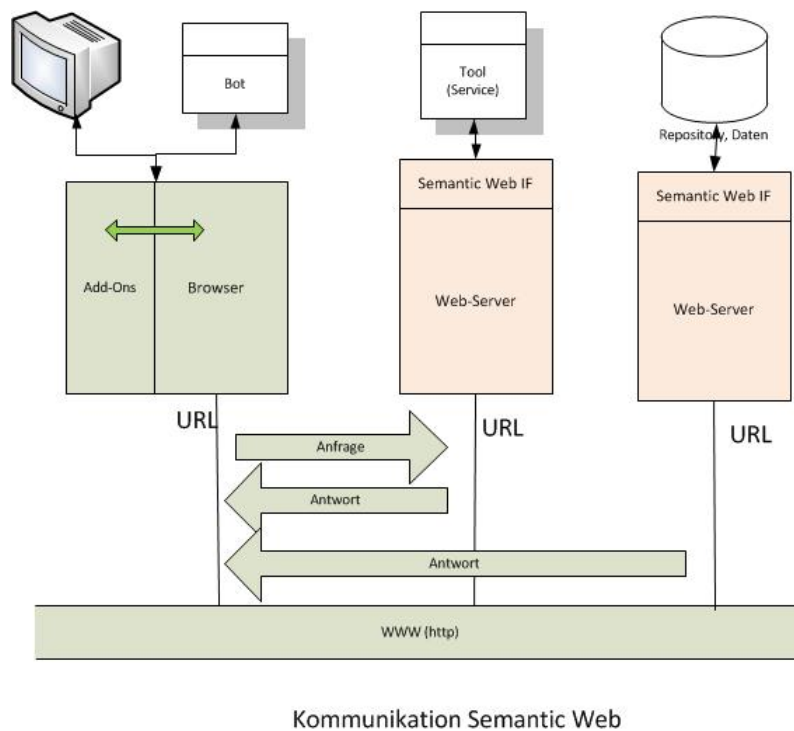


Abbildung 6.1: Kommunikationsequenz: Browser ist Endpunkt der Kommunikation

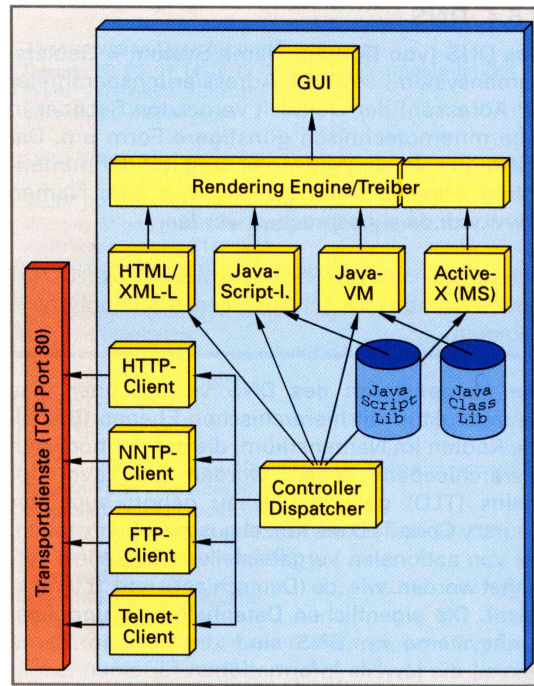


Bild 1: Aufbau eines Web-Browsers unter dem Betriebssystem Windows

Abbildung 6.2: Architektur eines Browser unter Windows

(Quelle Gretzinger, 2007, S. 490)

Die Kommunikation verläuft also zwischen Browsern¹⁶⁰ (Client) und Webservern. Sie wird mit dem WWW üblichen Protokoll-Standards abgewickelt. Das bedeutet, dass die Kommunikation http als Transportprotokoll (tcp/ip¹⁶¹ auf unterer Stufe) verwendet. Die Daten werden in http¹⁶² (meist HTML, XHTML, XML) codiert.¹⁶³

Diese Semantikmodellierung benötigt mindestens drei Funktionalitäten:

- Erstellung der Ontologie (Ontologie-Editor; Serverseitig)
- Abfragemechanismus (Query; Browser- und Serverseitig)

¹⁶⁰Als Bsp. einer Browserrealisierung siehe Fig. 6.2 S. 49.

¹⁶¹Siehe Fig. 6.3 S. 50.

¹⁶²Siehe Fig. 6.4 S. 51.

¹⁶³HTML, XHTML und XML sind Protokolle, die auf ASCII-Zeichen aufbauen (Text). So sind Tags, Attribute, Namen usw. als lesbarer Text ausgeführt. Der übertragenen Datenstrom kann auch codierte Daten oder Binärdaten (z.B.: Bilder) enthalten. Diese Datenabschnitte müssen aber in Anweisungen eingebettet sein, welche als Text den Datenstrom beschreiben.

- Visualisierung von Ontologieinformation (Browserseitig)

Zur Ontologie-Erstellung gibt es Editoren. Der Anwender kommt mit diesen Ontologien in Form von Referenzen in Kontakt. Der Datendistributor erstellt die Ontologieeinträge beim erstellen der Datendatei durch Eintragen der Referenzen, deren Cardinalität und Typs. Diese Verweise sind dann im Datenfile hinterlegt und werden bei Aktivierung durch den Browser abgefragt und aktualisiert. Dieses Abfragen nennt man „Harvesten“ (Ernten). Es wird dabei im Client das semantische Verweisnetzwerk aufgebaut.¹⁶⁴

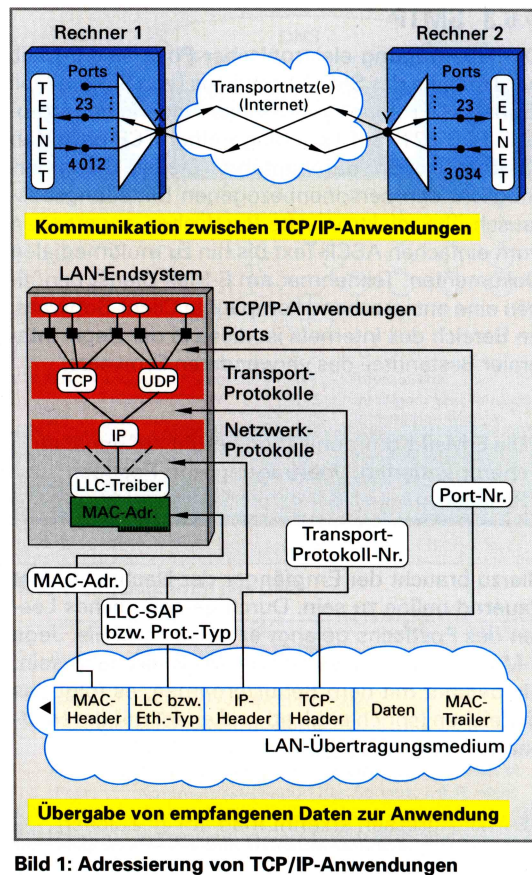


Bild 1: Adressierung von TCP/IP-Anwendungen

Abbildung 6.3: Anwendungsadressierung der TCP/IP Kommunikation

(Quelle Gretzinger, 2007, S. 487)

Spezifika, die für das „Semantic Web“ notwendig werden (zusätzliche Funktionalitäten), werden browser-seitig als Add-Ons realisiert, server-seitig als Services.

¹⁶⁴Als Beispiel zur Ontologiehandhabung siehe SIOC. (siehe SIOC, 2010).

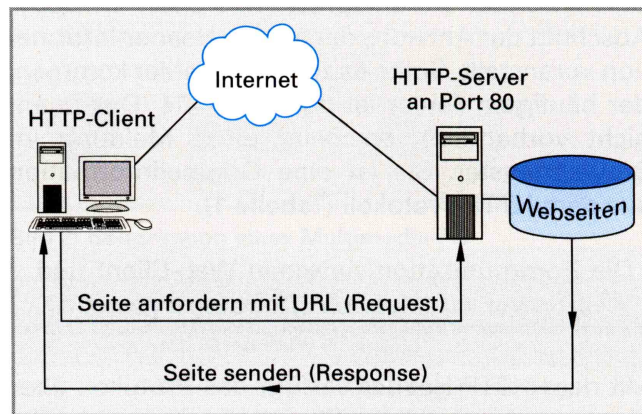


Bild 3: Ablauf einer vereinfachten HTTP-Sitzung

Abbildung 6.4: Schema einer http-Sitzung

(Quelle a. a. O., S. 489)

7 Vergleich von Realisierungsalternativen

Basierend auf dem Aufsatz *The Semantic Web*¹⁶⁵ von Berners-Lee soll versucht werden, eine Anforderungsbeschreibung mit DV-technischer Machbarkeitsüberlegung zu erstellen. Anhand dieser Überlegungen soll die architektonische und funktionale Begründung der für „Semantic Web“ spezifische WWW-Erweiterungen gezeigt werden.

7.1 Anwenderanforderungen

Als Anwendungsszenario schreibt Berners-Lee:

The entertainment system was belting out the Beatles' „We Can Work It Out“ when the phone rang. When Pete answered, his phone turned the sound down by sending a message to all the other local devices that had a volume control. His sister, Lucy, was on the line from the doctor's office: „Mom needs to see a specialist and then has to have a series of physical therapy sessions. Biweekly or something. I'm going to have my agent set up the appointments.“ Pete immediately agreed to share the chauffeuring.

At the doctor's office, Lucy instructed her Semantic Web agent through her handheld Web browser. The agent promptly retrieved information about Mom's prescribed treatment from the doctor's agent, looked up several lists of providers, and checked for the ones in-plan for Mom's insurance within a 20-mile radius of her home and with a rating of

¹⁶⁵Siehe Berners-Lee/Hendler/Lassila, 2001.

excellent or very good on trusted rating services. It then began trying to find a match between available appointment times (supplied by the agents of individual providers through their Web sites) and Pete's and Lucy's busy schedules. (The emphasized keywords indicate terms whose semantics, or meaning, were defined for the agent through the Semantic Web.)¹⁶⁶

So beginnt die Beschreibung des „Semantic Web“ bei Berners-Lee. Folgende Anforderungen (Requirements) lassen sich erkennen (siehe auch Fig. 7.1 S. 61):

1. Das Telefon versteht, dass das Telefongespräch durch die Stereoanlage gestört wird.
2. Abfrage von verschriebenen Dienste (Psychoterapeuten) basierend auf Öffnungszeiten unter Berücksichtigung von vorhandenen Terminen.
3. Abfrage von Psychoterapeuten im Umkreis von 25 KM die Kassenvertragsärzte sind.
4. Hinzuziehung von Ratinglisten der infrage kommenden Ärzte.
5. Vertrauenswürdiger Agent verändert Ergebnisliste eines anderen Agenten.

Eine etwas einfachere und klarere Anforderung ist bei Geisler zu finden. Er reduziert die Anforderungen an ein „Semantic Web“ klar auf die Verfolgung von Links. Die Browser-Add-Ons sollen die Übernahme gefundener Links in weitere Suchabfragen ermöglichen – also das Copy und Paste im Browser automatisieren.¹⁶⁷

In den nachfolgenden Überlegungen wird allerdings vorerst von den Visionen Berners-Lee ausgegangen. Wie das realisiert werden soll, stellt dieser sich wie folgt vor:

The Semantic Web will bring structure to the meaningful content of Web pages, creating an environment where software agents roaming

¹⁶⁶a. a. O.

¹⁶⁷Siehe Geisler, 2009, S. 11-14.

from page to page can readily carry out sophisticated tasks for users. Such an agent coming to the clinic's Web page will know not just that the page has keywords such as „treatment, medicine, physical, therapy“ (as might be encoded today) but also that Dr. Hartman works at this clinic on Mondays, Wednesdays and Fridays and that the script takes a date range in yyyy-mm-dd format and returns appointment times. And it will „know“ all this without needing artificial intelligence [...]. Instead these semantics were encoded into the Web page when the clinic's office manager (who never took Comp Sci 101) massaged it into shape using off-the-shelf software for writing Semantic Web pages along with resources listed on the Physical Therapy Association's site.¹⁶⁸

Diese Anforderung kann nun DV technisch unterschiedlich realisiert werden. Ob und warum zur Realisierung dieser Szenarien ein „Semantic Web“ als neue Qualität notwendig ist, soll die Analyse zeigen. Auf alle Fälle ist hier der klare Hinweis, dass die Struktur im Bedeutungsinhalt in die Webseiten eingebracht wird und eine Webumgebung zur Abfrage geschaffen wird.¹⁶⁹

7.2 Lösungsansätze

Es ist im Nachfolgenden zu beachten, dass immer mehrere Lösungen möglich wären – die gewählte also eine aus vielen möglichen ist. Mit den dargestellten Lösungen sollen die paradigmatischen Unterschiede in der Herangehensweisen dargestellt werden.

7.2.1 Konventionelle Lösung

Fall 1: Steurecode für Lautstärke (Individueller Befehlssatz eines Service)

Fall 2,3: Datenbank-Abfrage; Aufbereitung mittels Reportgenerator

¹⁶⁸Berners-Lee/Hendler/Lassila, 2001.

¹⁶⁹Dies ist im W3C-Ansatz mit den Konzepten RDF bis SPARQL verwirklicht.

Fall 4: Read/Write/Update-Interface (Bestandteil eines Datenbankinterface)

Im Allgemeinen ist der Lösungsansatz realisierbar.

Der Fall 1 wird mit gerätespezifischen Modulen gelöst, die im Telefon vom Hersteller als Browser oder Web-Bot realisiert sind und mit dem Nachrichtenteil der steueranlagenspezifisch ist vor Inbetriebnahme parametrieren muss (Anlagen-IP; Steuerbefehl für „Stumm“ und Steuerbefehl für „Musik“ in Abhängigkeit von den Telefonereignissen „Anruf kommt“ und „Gesprächsende“.)

Der Fall 2 und 3 muss durch Abfragen und Datenviews erstellt werden. Dazu muss die abzufragende Datenbank bekannt sein. Eine Assistant-Unterstützung zur Abfrageerstellung ist machbar. Die Parametrierung des Reportgenerators zur Abfrageanzeige ist ebenfalls mit Assistant-Unterstützung möglich. Abfrage-Editor und Report-Assistent wären als Anwenderprogramm am PC zu installieren (etwa in der Art wie Citavi¹⁷⁰ oder ähnliche Programme).

Der Fall 4 ist nur mit weiteren Zusatzprogrammen zu realisieren. Eine Kommunikation mit dem Agenten scheint nur über Mail möglich, wenn nicht ein Webserver am Ziel-PC installiert oder andere Zugriffe auf den PC erlaubt werden sollen (etwa in der Art wie SW-Updates laufen).

Damit ergibt sich folgendes Bild:

Vorteil: Sofort realisierbar und auch schon realisiert (z.B. Google-Suche, SW-Update, Bibliotheksabfragen wie Citavi usw.).

Nachteil: Teilweise gerätespezifisch; geschlossene Systeme

Entscheidender Nachteil: Die Datenmenge die im WWW vorhanden ist, kann nur punktuell genutzt werden!

Zwar ist es mit Konzepten, die Dienstrepositories anbieten möglich, nicht auf die Informationsmenge eines einzigen Anbieters zurückgeworfen zu sein - aber die nutzbare Information ist trotz dessen nur ein Bruchteil der im Web vorhandenen. Zudem dürfte in dieser Variante die Informationsbereitstellung nicht kostenfrei

¹⁷⁰Citavi ist ein Literaturverwaltungs- und Wissensorganisationsprogramm. (Citavi, 2011)

sein.¹⁷¹

Der entscheidende Unterschied ist das, was Hitzler „Interoperabilität“ nennt: Die serviceübergreifende Kommunikationsverarbeitung ist nicht gewährleistet. Die Datenverarbeitung ist dienstanbieterspezifisch.

Will man diesen Ansatz generalisieren, so erfordert das Festlegungen: Vor allem für die Datenbankschnittstelle muss eine Abfragesprache definiert werden (SQL oder Teilmenge von SQL). Die Datenbankergebnisse müssen mit dem Reportgenerator harmonieren und beide müssen im Browser implementiert sein (Aufrufbar sein). Aber vor allem: Die Datenquellen müssen bei Abfragebeginn bekannt sein. Für unbekannte Abfragequellen dürfte die Realisierung schwierig sein. Es müsste dann auf ein SOA-ähnliches Konzept ausgewichen werden – etwa einer CORBA-Lösung mit Service-Broker (Namensdienst). Aber auch dann ist es kaum möglich, die bereits vorhandene Informationsmenge im WWW sinnvoll zu nutzen.¹⁷²

All diese Änderungen und Abmachungen sind es, die zum „Semantic Web“ – Ansatz geführt haben.

An dieser Stelle noch ein vergleichender Hinweis zum Thema Datenbank versus Ontologie (siehe Fig. 7.2 S. 62). Worin ist der große Unterschied? Es gibt zwei:

1. Datenbanken benötigen ein Schema bevor sie Daten aufnehmen können
2. Datenbanken müssen konsistent sein

Das Datenbankschema beschreibt, welche Tabellen und Views die Datenbank umfasst. Es wird also ein hierarchischer Tabellenbaum spezifiziert, in dem die Datenbestände möglichst Redundanzfrei¹⁷³ erfasst werden.

¹⁷¹Die Konzepte die auf SOA aufbauen sehen Servicebroker vor, welche Dienstanbieter auflisten. Es bietet sich an, dass der Zugang vergibt wird. Auch Microsoft verfolgte und verfolgt den Ansatz, Dienste auf diese Weise anzubieten.

¹⁷²Die Nutzung der vorhandenen Informationsmenge ist ein entscheidender Grund für das „Semantic Web. Diese Nutzung wird durch offene Protokolldefinitionen erreicht (RDF).“ (vgl. Schlönvoigt, 2010) Das Problem, vorhandene Quellen nicht nutzen zu können, haben aber auch manche Lösungsansätze für das „Semantic Web“. Vor allem, wenn Security in die Betrachtungen eingebracht wird.

¹⁷³Redundante Daten in einer Datenbank gelten als pathologisch, da sie zu Updateproblemen und Inkonsistenzen führen. Datenbanken werden im allgemeinen nach den sogenannten Normalformen – also keine Datenredundanz und keine referentielle Redundanz – entworfen. (Als Übersicht Zilahi-Szabó, 2002) Redundanzen sind nur in Ausnahmefällen aus Performancegründen erlaubt.

7.2.2 Lösung als SOA (CORBA)

SOA heißt serviceorientierte Architektur und bedeutet, dass Softwarefunktionalität als genau definierte Dienstleistung angeboten wird. SOA ist ein Architekturmodell, ein abstraktes Konzept, das festlegt, welche Funktionalitäten ein SOA - Framework enthalten muss, um Services beliebiger Granularität sinnvoll anbieten zu können.

Gleichzeitig ist SOA auch ein Beispiel, wie man solche Probleme nach dem „alten Paradigmen“ angeht: Möglichst umfassende Analyse - möglichst eindeutige Spezifikation und Begriffsbildung. Das Ergebnis sind komplexe Modelle, die dann schlecht verstanden werden und nicht zu den geplanten Ergebnissen führen. Dies mag eine der Gründe sein, dass Berners-Lee gleich auf die durchgängige Modellierung verzichtet und alles der Community überlässt.

Fast jede größere Softwarefirma hat ein eigenes SOA-Modell.¹⁷⁴ Auch Open Source SOA-Stack sind verfügbar.¹⁷⁵

Zum Vergleich sollen nachfolgend einige Basiselemente von SOA dargestellt werden.¹⁷⁶

Die Kernparadigmen der SOA sind als „SOA Tempel“ bekannt ((siehe Fig. 7.4 S. 63).

Wie innerhalb einer SOA kommuniziert wird, ist ebenfalls klar geregelt (siehe Fig. 7.5 S. 63).

Jeder Dienst muss in einer SOA-Umgebung zuerst von einem Anbieter in einem Dienstverzeichnis veröffentlicht werden. Dienstanwender fragen immer solche Verzeichnis ab, wählen einen Dienst aus und werden so mit dem Dienstanbieter verbunden.

Vorhandene WWW-Informationen kann nur sehr eingeschränkt nutzbar gemacht werden. Es können nur Dienste und Dienstbeschreibungen ausgewählt werden, die, ähnlich einem Branchenverzeichnis, als Beschreibungslisten registriert sind. Dies ist aber keine Einschränkung. Der Namensdienst einer SOA-Realisation ist prinzipiell in der Lage, weltweit Zugang zu ermöglichen.¹⁷⁷ Eine SOA-Lösung kann sich also

¹⁷⁴Siehe Liebhart, 2007, S. 39-58.

¹⁷⁵Siehe a. a. O., S. 56.

¹⁷⁶Als Beispiel eine SOA-Architektur siehe Fig. 7.3 62.

¹⁷⁷Man denke nur an den DNS des tcp/ip Routings oder an den Domainservice von Windows-Domains.

beliebig nahe dem Funktionsumfang der Vorstellung Berners-Lee nähern. Darum werden in dem Buch *Service-orientierte Architekturen mit Web Services* die Semantic-Web-Technologie als *Semantik Web Services*¹⁷⁸ dargestellt.

Der jeweilige Service wird mittels WSDL¹⁷⁹ beschrieben.¹⁸⁰ Abfragen auf definierte QoS-Parameter¹⁸¹ sind möglich. Es gibt die Möglichkeit, Prozessabläufe durch Orchestrierung zu gestalten. Zur Prozessgestaltung kann BPEL¹⁸² verwendet werden.¹⁸³

Die SOA - Spezifikation ist durchgängig definiert. So ist auch die Sichtweise, unter denen man SOA betrachten kann im Modell enthalten. Die unterschiedlichen Betrachtungsweisen und ihr Zusammenhang sind als Aspekt modelliert.¹⁸⁴

Ebenso ist der Web-Service-Stack¹⁸⁵ festgelegt.¹⁸⁶

Die einzelnen Teilmodelle

- Nachrichten-Modell
- Service-Modell
- Ressourcen-Modell
- Richtlinien-Modell

sind ebenfalls vollständig modelliert. Als Beispiel sei hier das Service-Modell gezeigt (siehe Fig. 7.8 S. 65).

Mit SOA ist „Semantic Web“ also prinzipiell machbar, da „Semantic Web“ eine verteilte Service Architektur hat. SOA zielt auf die Modellierung verteilter Anwendungen und nicht, wie „Semantic Web“ auf die Abfrage verteilter Daten. SOA enthält somit viele, für „Semantic Web“ brauchbare Konzeptansätze und Festlegungen, ist aber für das Abfragen verteilter Daten nicht konzipiert.

¹⁷⁸Dostal et al., 2005, S. 287.

¹⁷⁹Web Service Description Language

¹⁸⁰Vgl. Liebhart, 2007, S. 16-17

¹⁸¹QoS = Quality of Service: Verfügbarkeit, Performance u.ä. Dienstparameter.

¹⁸²BPEL = Business Process Execution Language

¹⁸³Vgl. Liebhart, 2007, S. 18-22

¹⁸⁴Siehe Fig. 7.6 S. 64.

¹⁸⁵Schichtenmodell der Dienstkommunikation

¹⁸⁶Siehe Fig. 7.7 S. 64.

Eine konkrete Spezifikation für verteilte Dienste ist die der OMG¹⁸⁷. Für die Anwendung braucht man noch weitere Spezifikationen der zu installierenden Komponenten. Für die SOA der OMG würde eine mögliche Spezifikation der Realisierung CORBA sein. Die tatsächlich realisierten Dienstumgebungen können dann herstellerspezifische Softwarepakete wie z.B. *Distributed Smalltalk*¹⁸⁸, *VisiBroker*¹⁸⁹ oder *DSOM*¹⁹⁰ sein.¹⁹¹ SOA - Realisationen, die sich nicht an die OMG Spezifikationen halten - also zwar SOA - Architekturen aber keine CORBA realisierungen sind, gibt es auch. DCOM/OLE¹⁹² von Microsoft sei hier stellvertretend genannt.¹⁹³

Lösung mit CORBA:

CORBA ist eine Spezifikation einer Infrastruktur für eine serviceorientierte Architektur.¹⁹⁴ Sie wird herstellerspezifisch realisiert. Die Realisationen der einzelnen Hersteller sind weitgehend schnittstellenkompatibel. Für eine Lösung der gestellten Aufgabe ist also Voraussetzung, dass die benötigten Dienste durch eine CORBA-Realisation angeboten werden. In dieser Umgebung veröffentlichen dann Dienstanbieter ihre Dienste. Die Nutzer (Clients) werden durch anwendungsspezifische Software realisiert, welche die benötigten Dienste dann von einem ORB (Object Request Broker) erfragen. Die externen Zugriffe erfolgen mittels RPC¹⁹⁵. Die Dienstschnittstellen sind in der Clientanwendung mittels Proxies¹⁹⁶ realisiert. Die Datengewinnung und Auswertung kann mittels Orchestrierung modelliert werden.¹⁹⁷

Vorteil: Sofort realisierbar und auch schon realisiert. Stabile klare Softwarestruktur, kommerzielle saubere „Dienstnutzung“.

Nachteil: Herstellerabhängig; kein offenes Konzept; Client-Serverkommunikation

¹⁸⁷OMG = Object Management Group.

¹⁸⁸Eine CORBA-Plattform von ParcPlace Digital

¹⁸⁹Eine CORBA-Plattform von Visigenic

¹⁹⁰DSOM = Distributed System Object Model der Firma IBM

¹⁹¹Vgl. Sayegh, 1997, S.16-17

¹⁹²DCOM = Distributed Component Object Model; OLE = Object Linking & Embedding

¹⁹³Vgl. Sayegh, 1997, S. 15.

¹⁹⁴OMG, 2008, Vgl..

¹⁹⁵RPC = Remote Procedure Calls.

¹⁹⁶Proxies sind Platzhalter, welche die Schnittstellen der entfernten Anwendung zum Zeitpunkt der Softwarecompilierung zur Verfügung stellen.

¹⁹⁷Vgl. Liebhart, 2007, S. 83-89.

mit RPC¹⁹⁸.

Entscheidender Nachteil: Die Datenmenge die im WWW vorhanden ist, kann nur punktuell genutzt werden!

Die Lösung soll hier nicht weiter ausgeführt werden, da zu erkennen ist, dass diese Art der Realisierung den Paradigmen der Offenheit des WWW klar widerspricht. Auch die Art der Dienstnutzung durch Clientanwendungen die aufgabenspezifisch erstellt werden müssen und RPC¹⁹⁹ enthalten, entspricht nicht dem Gedanken des WWW.

¹⁹⁸RPC = Remote Procedure Call. Die Clientprogramme enthalten Prozeduraufrufe an den Server. Die Kommunikation erfolgt nicht datenbasiert.

¹⁹⁹RPC = Remote Procedure Calls

Anwendungsfälle Semantic Web

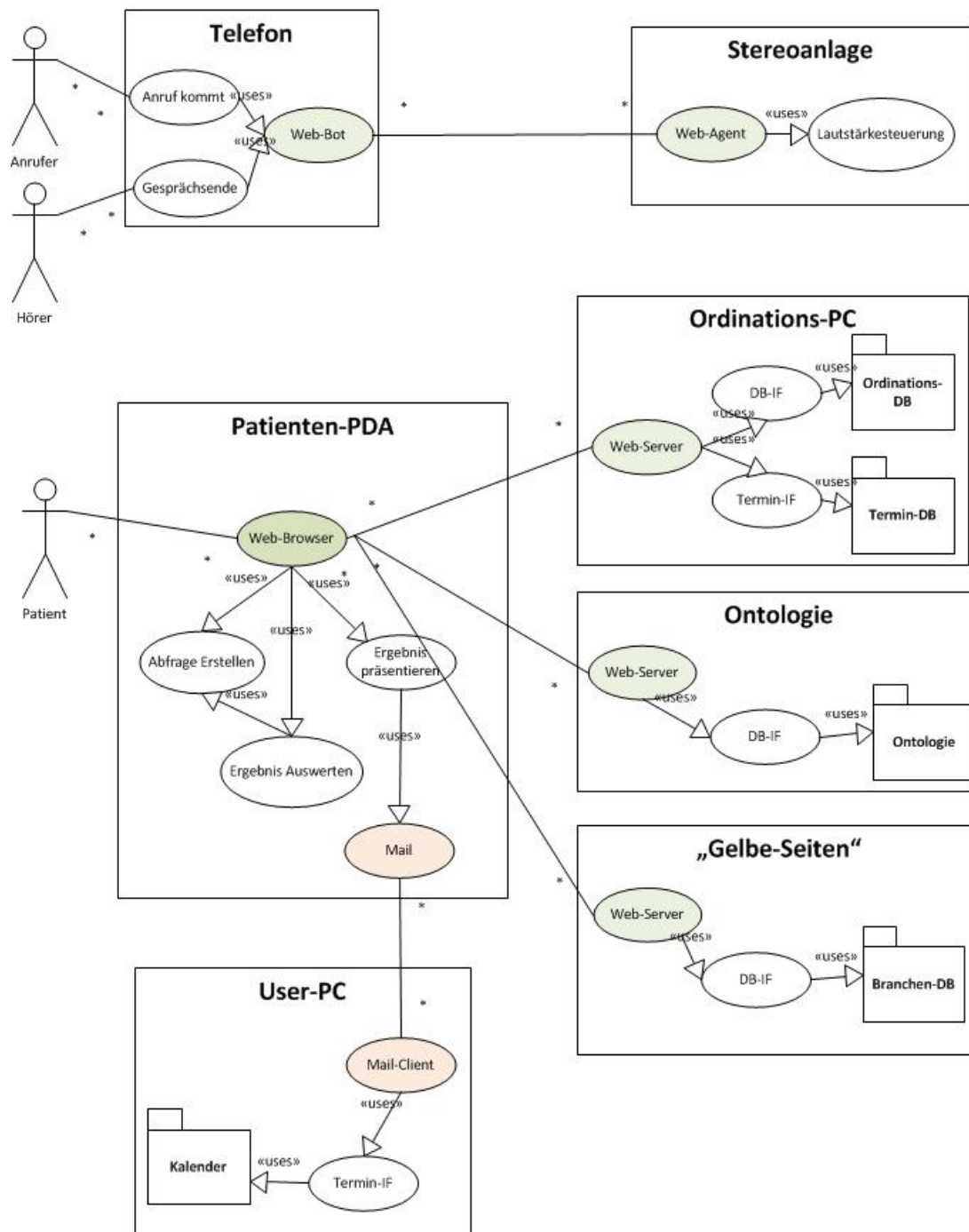


Abbildung 7.1: Anwendungsfälle für Semantic Web

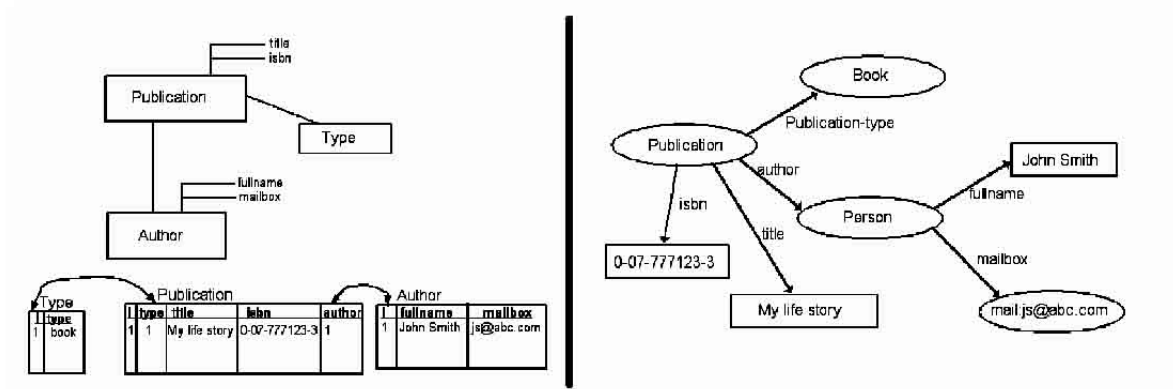


Fig. 8. An entity relationship diagram with a database implementation (left) and a RDF diagram (right) modelling the *publication* data of the scenario.

Abbildung 7.2: Unterschied Relationale DB vs. Ontologie

(Quelle Gerber/van der Merwe/Barnard, 2007, S. 12)

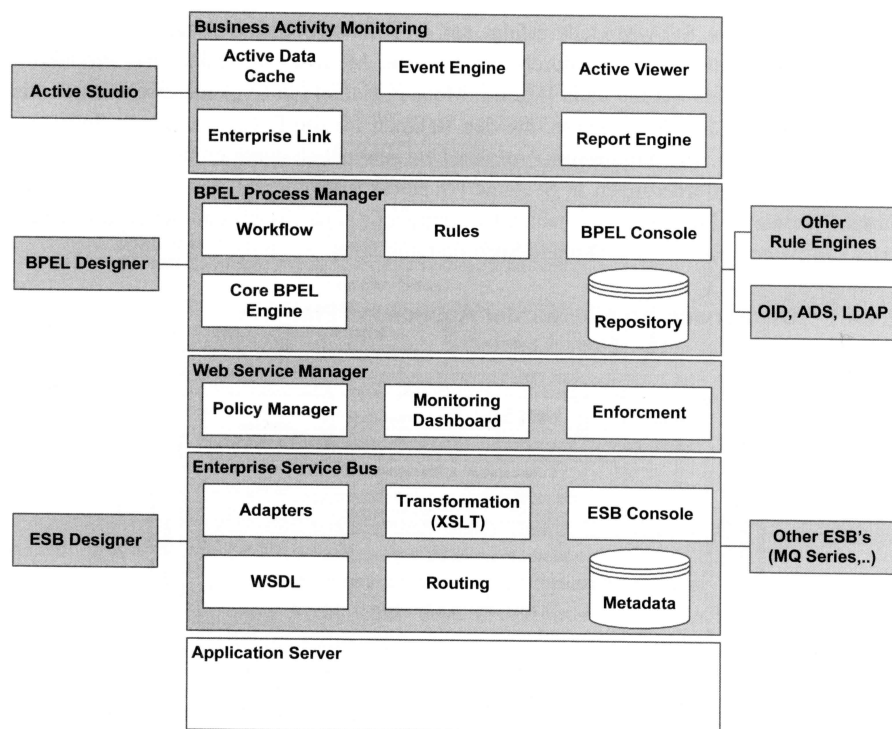


Abbildung 2.17 Die Komponenten der Oracle SOA-Suite

Abbildung 7.3: Oracle SOA-Suite

(Quelle Liebhart, 2007, S. 51)

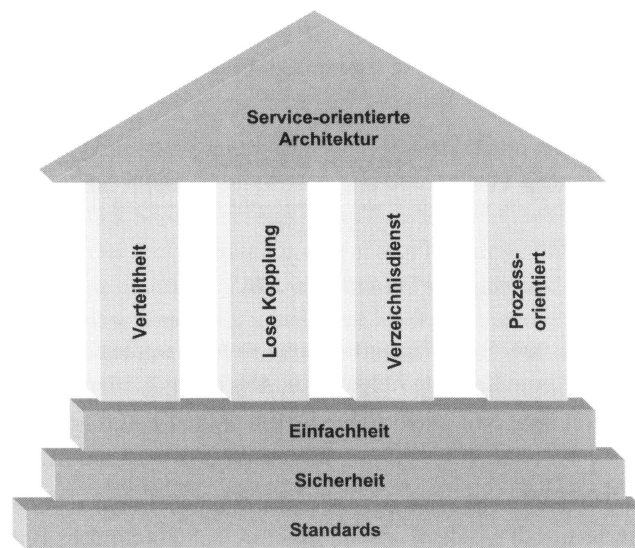


Abbildung 2.1 SOA Tempel

Abbildung 7.4: SOA - Entwurfsparadigmen

(Quelle: Dostal et al., 2005, S. 11)

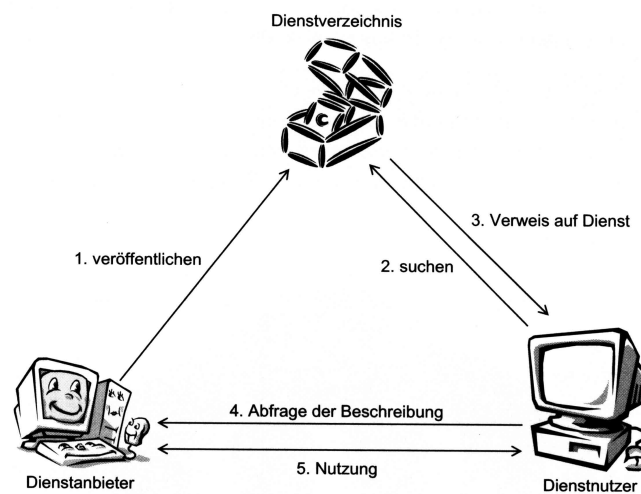


Abbildung 2.2 Das magische Dreieck einer SOA

Abbildung 7.5: Kommunikationsbeziehungen SOA-Komponenten

(Quelle: a. a. O., S. 12)

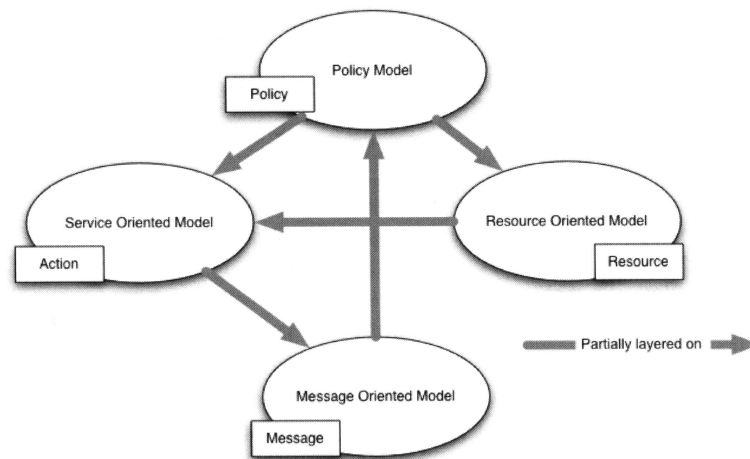


Abbildung 3.3 Aspekte und Abhängigkeiten einer Web-Services-Architektur

Abbildung 7.6: Aspekte einer SOA

(Quelle: Dostal et al., 2005, S. 31)

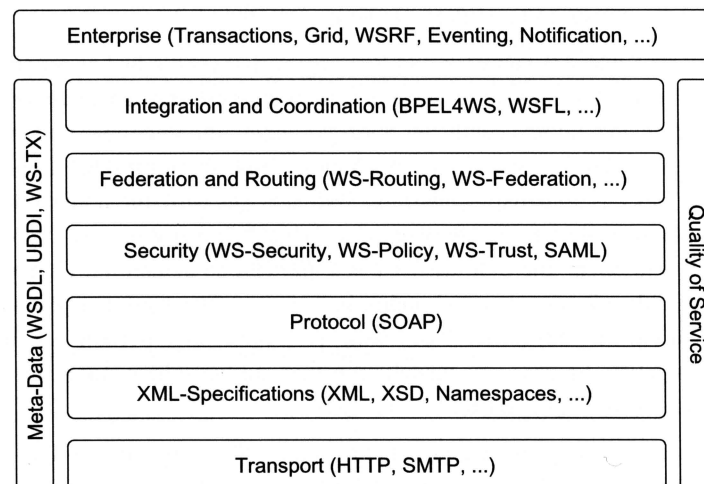


Abbildung 3.2 Web Services Stack

Abbildung 7.7: Schichtenmodell der Dienste

(Quelle: Dostal et al., 2005, S. 29)

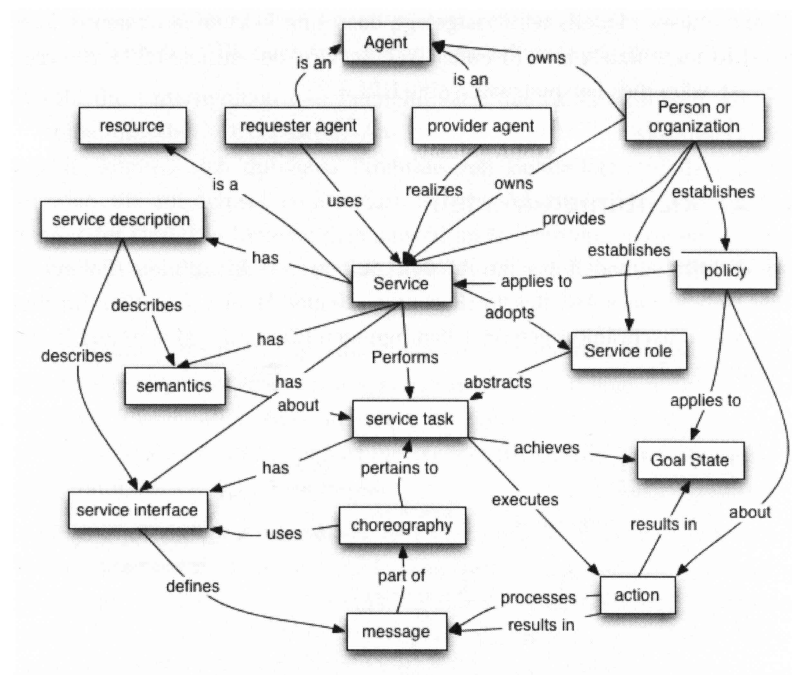


Abbildung 3.5 Architekturmodell eines Services von Web Services

Abbildung 7.8: SOA Services-Modell

(Quelle: a. a. O., S. 33)

8 Resümee

Die funktionale Vision von Berners-Lee kann – speziell durch SOA – realisiert werden (und ist im industriellen Bereich auch im Einsatz). Die paradigmatische Idee der offenen Kommunikation ist in diese Realisation nicht enthalten.

Es ist daher berechtigt, die Vision Berners-Lee als einen konsequenten Paradigmenwechsel mit ungeahntem Potential zu sehen. Offene Kommunikation, verteilte, individualisierte Lösungsentwicklung basierend auf einem evolutionär Konzept mit Potential für Emergenzen ist ein Weg, der verfolgt werden sollte. Der Ansatz, komplexe Probleme in einfache Einzelbausteine aufzulösen und diese im Rahmen einer gemeinsamen Ziel-Vision durch die Web-Community Lösungen zuzuführen kann zu kreativen Resultaten führen. Allerdings sind die „Kosten“ dieses Weges der Verlust von Systematik und Durchgängigkeit. Diese Zunahme an „Chaos“ kann aber eben der Faktor sein, der die Probleme, an denen die KI²⁰⁰ bisher gescheitert ist, löst.

Zu beachten ist auch, dass es eben „das“ „Semantic Web“ noch nicht gibt. So versucht jeder, möglichst kraft seiner Reputation, seinen Standpunkt als den zentralen zu positionieren und so die Entwicklung in eine bestimmte Entwicklung zu treiben.²⁰¹ Es ist daher entscheidend, „seinen“ Interessensschwerpunkt immer im Kopf zu behalten – was aufgrund der vielen Einflüsse oft schwer ist.

Mir ist es in dieser Reflexion gelungen, mich ein bisschen in die Thematik einzuarbeiten. Leider bleibt noch immer vieles verschlossen. Die Nachteile der pluralistisch, anarchischen, aufgabenkomplexitätsreduzierenden Vorgehensweise kommen bei mir voll zum Tragen. Das Feld des WWW ist inzwischen so vielfältig gefüllt mit Lösungen, Teillösungen, Vorschlägen und anderer Elemente, dass die Komplexität erheblich ist. Es gibt eine Unmenge von Beschreibungen und Darstellungen,

²⁰⁰KI = Künstliche Intelligenz

²⁰¹Siehe als Bsp.: Kingsley, 2010.

die in vertretbarer Zeit kaum zu bearbeiten sind. Semantic Web setzt auf HTML, XML, XPath, RDF usw. usw. auf. Für jedes dieser Kürzel existieren Folianten mit mehreren 100 Seiten. Aus meiner Sicht ist es nicht klar, ob das Verlassen des klassischen Weges ein Problem kanonisch zu realisieren richtig ist. Mir geht es jedenfalls so, dass das Erlernen angeblicher komplizierter Sprachen wie C oder C++ im Verhältnis zum Geflecht der „einfachen“ Web-Konzepte ein Kinderspiel war. Bei mir halten sich immer noch Reste der Einstellung, ein effizientes Datenübertragungsprotokoll und leistungsstarke Compiler zur Erstellung der funktionalen Applikationen könnte mehr Nutzen bringen – andererseits ist die angerissene Problematik auch sehr reizvoll und verführerisch.

Literaturverzeichnis

- Berners-Lee, Tim/Hendler, James/Lassila, Ora (2001):** The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. Scientific American, Nr. May $\langle$ URL: <http://www.scientificamerican.com/article.cfm?id=the-semantic-web> $\rangle$ – Zugriff am 12.3.2011.
- Dose, Maria (cop 1990):** Duden - Fremdwörterbuch. Band Band 5, Ed. 5, Der Duden in 12 Bänden, 5. Auflage. Mannheim: Dudenverlag, ISBN 3-411-20915-1.
- Dostal, Wolfgang et al. (2005):** Service-orientierte Architekturen mit Web Services: Konzepte - Standard - Praxis. München: Elsevier Spektrum Akademischer Verlag.
- Eco, Umberto/Trabant, Jürgen (1994):** Einführung in die Semiotik. Band 105, UTB für Wissenschaft;Linguistik, Literaturwissenschaft, Philosophie, Autorisierte deutsche Ausg., 8., unveränd. Aufl. Auflage. München: Wilhelm Fink Verlag, ISBN 3-7705-0633-2.
- Føllesdal, Dagfinn (1980):** Semantik. In **Speck, Josef (Hrsg.):** Handbuch Wissenschaftstheoretischer Begriffe. Göttingen: Vandenhoeck & Ruprecht, S. 569-579.
- Geisler, Matthias (2009):** Semantic Web. Unterhaching: Entwickler.press, ISBN 978-3-86802-028-1.
- Gretzinger, Klaus (2007):** IT-Fachkunde. 3. Auflage. Haan-Gruiten: Verl. Europa-Lehrmittel, ISBN 978-3-8085-3653-7.

- Heyl, Andreas (2005):** Nichtlineare Dynamik und Komplexität in Produktionssystemen: Zugl.: Stuttgart, Univ., Diss. 2005 (Nicht für den Austausch). Berlin: dissertation.de - Verlag im Internet, ISBN 3-86624-068-6.
- Hitzler, Pascal et al. (2008):** Semantic Web: Grundlagen. eXamen.press, Berlin: Springer-Verlag, ISBN 978-3-540-33993-9.
- Jcobi, Ingeborg (1995):** Softwaretechnik und Programmiersprachen. In **Werner, Dieter (Hrsg.):** Taschenbuch der Informatik. Leipzig: Fachbuchverlag Leipzig GmbH, S. 324-392.
- Liebhart, Daniel (2007):** SOA goes real: Service-Orientierte Architekturen erfolgreich planen und einführen. München: Carl Hanser Verlag.
- Luhmann, Niklas (1987):** Soziale Systeme: Grundriß einer allgemeinen Theorie. Frankfurt am Main: Suhrkamp.
- Meinhardt, Jürgen (1995):** Kap. 1 Grundbegriffe: 1.5.2 Frühzeit der Computerentwicklung. In **Werner, Dieter (Hrsg.):** Taschenbuch der Informatik. Leipzig: Fachbuchverlag Leipzig GmbH, S. 19-27.
- Messinger, Heinz/Rüdenberg, Werner (1974):** Langenscheidts Handwörterbuch Englisch: Teil I Englisch-Deutsch. 17. Auflage. Berlin: Langenscheidt.
- Orfali, Robert/Harkey, Dan/Edwards, Jeri (1998):** Instant CORBA: Führung durch die CORBA-Welt. Bonn u.a.: Addison-Wesley, ISBN 3-8273-1325-2.
- Petrasch, Roland/Fieber, Florian/Bednarz, Maciej A. (2008):** Agile modellbasierte Softwareentwicklung in der Praxis. OBJEKTSpektrum, Nr. Nr. 3, S. 36-44.
- Sayegh, Andreas (1997):** CORBA: Standard, Spezifikationen, Entwicklung. Cambridge: O'Reilly Verlag.
- Schiller, Jennifer/Canditt, Sabine (2008):** Das Ganze sehen: Auf dem Weg zum agilen Unternehmen. OBJEKTSpektrum, Nr. Nr. 3, S. 24-28.

Speck, Josef (Hrsg.) (1980): Handbuch Wissenschaftstheoretischer Begriffe:
Band 3 (R.Z). Göttingen: Vandenhoeck & Ruprecht.

Türker, Can/Saake, Gunter (2006): Objektrelationale Datenbanken: Ein
Lehrbuch. Heidelberg Neckar: dpunkt, ISBN 3-89864-190-2.

Internetquellen

Citavi (2011): Citavi: Literaturverwaltung & Wissensorganisation. $\langle$ URL: <http://www.citavi.com/> $\rangle$ – Zugriff am 3.4.2011.

DAML (2010): OWL for Services (OWL-S). $\langle$ URL: <http://www.ai.sri.com/daml/services/owl-s/> $\rangle$ – Zugriff am 14.7.2011.

Dublin Core Metadata Initiative (2010): Dublin Core Metadata Element Set, Version 1.1. $\langle$ URL: <http://dublincore.org/documents/dces/> $\rangle$ – Zugriff am 17.7.2011.

Fillies, Christian/Weichhardt, Frauke (2011): Graphische Entwicklung und Nutzung von Ontologien mit SemTalk in MS Office. $\langle$ URL: <http://subs.emis.de/LNI/Proceedings/Proceedings14/GI-Proceedings.14-11.pdf> $\rangle$.

Gerber, Aurna/Merwe, Alta van der/Barnard, Andries (2007): A Functional Semantic Web Architecture. $\langle$ URL: <http://ksg.meraka.org.za/~agerber/Paper152.pdf> $\rangle$ – Zugriff am 1.4.2011.

HB (2007): W3C will mit GRDDL Mikroformate und das Semantic Web verbinden. $\langle$ URL: <http://www.heise.de/ix/meldung/W3C-will-mit-GRDDL-Mikroformate-und-das-Semantic-Web-verbinden-174183.html> $\rangle$ – Zugriff am 14.7.2011.

Herman, Ivan (2007): Web Ontology Language (OWL). $\langle$ URL: <http://www.w3.org/2004/OWL/> $\rangle$ – Zugriff am 14.7.2011.

Herman, Ivan (2009): W3C Semantic Web Frequently Asked Questions. $\langle$ URL: <http://www.w3.org/2001/sw/SW-FAQ#relXML> $\rangle$ – Zugriff am 14.7.2011.

Jaenicke, Nadine (2009): Triplify. $\langle$ URL: <http://triplify.org/Overview> $\rangle$ – Zugriff am 13.7.2011.

Kingsley, Idehen (2010): 7 questions to Kingsley Idehen: "By declaring its context, Linked Data can be made more easily reusable by others". $\langle$ URL: <http://www.semantic-web.at/1.36.resource.308.7-questions-to-kingsley-idehen-x22-by-declaring-its-context-linked-data-made-more-easy-to-use.html> $\rangle$ – Zugriff am 3.4.2011.

Lena (2011): Catalog Of OMG Specifications. $\langle$ URL: http://www.omg.org/technology/documents/spec_catalog.htm $\rangle$ – Zugriff am 2.4.2011.

OMG (2008): Common Object Request Broker Architecture (CORBA) Specification, Version 3.1: Part 2: CORBA Interoperability: OMG Available Specification with change bars. $\langle$ URL: <http://www.omg.org/spec/CORBA/3.1/Interoperability/PDF/> $\rangle$ – Zugriff am 3.4.2011.

OpenLink (2010a): RDF Graf Model Engine. $\langle$ URL: <http://virtuoso.openlinksw.com/rdf-quad-store/> $\rangle$ – Zugriff am 9.7.2011.

OpenLink (2010b): Virtuoso Universal Server. $\langle$ URL: <http://virtuoso.openlinksw.com/> $\rangle$ – Zugriff am 9.7.2011.

Schlönvoigt, Steffen (2010): Semantic Web Technologies: SS 2010 - 01 - Einführung. $\langle$ URL: <http://www.slideshare.net/seso81/semantic-web-technologies-ss-2010-01-einfuehrung> $\rangle$ – Zugriff am 3.4.2011.

Schwartz, Aaron (2001): SWAG - Semantic Web Agreement Group. $\langle$ URL: <http://swag.webns.net/> $\rangle$ – Zugriff am 2.4.2011.

Semantic Web (2007): Rdfs:subClassOf. $\langle$ URL: <http://semanticweb.org/wiki/Rdfs:subClassOf> $\rangle$ – Zugriff am 13.7.2011.

SIOC (2010): SIOC Core Ontology Specification. $\langle$ URL: <http://rdfs.org/sioc/spec/> $\rangle$ – Zugriff am 18.7.2011.

- Stohlmann, Dirk (2006):** HiPath SESAP V1.0 Secured Enterprise Service and Administration Platform: Kundenpräsentation. [⟨URL: http://www.siemens-consultsupport.com/cms/images/downloads/portfolio/sesap/hipath_sesap_v1_0_kundenpraesentation.pdf⟩](http://www.siemens-consultsupport.com/cms/images/downloads/portfolio/sesap/hipath_sesap_v1_0_kundenpraesentation.pdf) – Zugriff am 19.3.2011.
- Sutter, Christoph (2007):** Certificat: Security Qualification for the Remote Service based on SIRA. [⟨URL: http://www.tuvit.de/certuvit/pdf/9532UE.pdf⟩](http://www.tuvit.de/certuvit/pdf/9532UE.pdf) – Zugriff am 19.3.2011.
- TEI (2011):** TEI: Text Encoding Initiative. [⟨URL: http://www.tei-c.org/index.xml⟩](http://www.tei-c.org/index.xml) – Zugriff am 15.7.2011.
- THESEUS (2010):** Das Theseus Forschungsprogramm: Neue Technologien für das Internet der Dienste. [⟨URL: http://theseus-programm.de/documents/BMWi_THESEUS_Broschuere_0510_D.pdf⟩](http://theseus-programm.de/documents/BMWi_THESEUS_Broschuere_0510_D.pdf) – Zugriff am 9.7.2011.
- Trunk, Daniela (2005):** Semantische Netze in Informationssystemen: Verbesserung der Suche durch Interaktion und Visualisierung. [⟨URL: http://www.fbi.fh-koeln.de/institut/papers/kabi/volltexte/band051.pdf⟩](http://www.fbi.fh-koeln.de/institut/papers/kabi/volltexte/band051.pdf) – Zugriff am 10.7.2011.
- Vetter, Johannes (2006):** Multimedia-Metadaten und ihre Anwendung: Semantic Web im Überblick. [⟨URL: http://www.medien.ifi.lmu.de/fileadmin/mimuc/hs_ws0506/presentations/Semantic_Web_Ueberblick.pdf⟩](http://www.medien.ifi.lmu.de/fileadmin/mimuc/hs_ws0506/presentations/Semantic_Web_Ueberblick.pdf) – Zugriff am 3.4.2011.
- W3C (2004):** RDF Primer. [⟨URL: http://www.w3.org/TR/2004/REC-rdf-primer-20040210/⟩](http://www.w3.org/TR/2004/REC-rdf-primer-20040210/) – Zugriff am 15.7.2011.
- W3C (2007):** SPARQL Query Language for RDF. [⟨URL: \label{fig:SOAServiceStack}⟩](#) – Zugriff am 14.7.2011.
- W3C (2010):** Semantic Web. [⟨URL: http://www.w3.org/standards/semanticweb/⟩](http://www.w3.org/standards/semanticweb/).

Wikipedia (2011): HL7. $\langle$ URL: <http://de.wikipedia.org/wiki/HL7> $\rangle$ – Zugriff am 9.7.2011.

Zilahi-Szabó, Miklós Géza (2002): Datenbank - Datenbanksysteme: Seminar 2001/2002. $\langle$ URL: http://docs.google.com/viewer?a=v&q=cache:EFvOVmoCH9QJ:www.informatik.uni-giessen.de/Sommersemester/SE/Datenbank_Normalisierung.ppt+normalformen+datenbank&hl=de&gl=at&pid=bl&srcid=ADGEESihNynu8fcfV_28VgMu4d7RWAnofn6GujwvPsNGYX066IGSSN6NE4RoCQuEGj28cPaD21GH0mxD6ac52v0gyBy1xiDLLy64fCA_E&sig=AHIEtbRdg8W7PpxIqy3JwznTdgW8MUrf3A $\rangle$ – Zugriff am 3.4.2011.