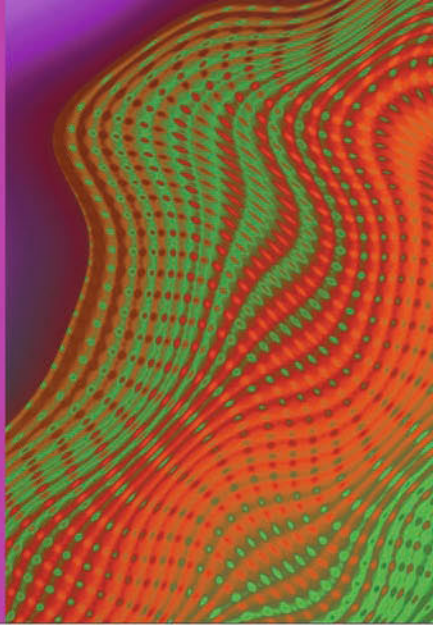




X . media . press



**Börteçin Ege
Bernhard Humm
Anatol Reibold (Hrsg.)**

X.media.press ist eine praxisorientierte Reihe zur Gestaltung und Produktion von Multimedia-Projekten sowie von Digital- und Printmedien.

Corporate Semantic Web

Wie semantische Anwendungen
in Unternehmen Nutzen stiften



Springer Vieweg

X.media.press

Weitere Bände in dieser Reihe
<http://www.springer.com/series/4332>

X.media.press ist eine praxisorientierte Reihe zur Gestaltung und Produktion von Multimedia-Projekten sowie von Digital- und Printmedien.

Börteçin Ege • Bernhard Humm • Anatol Reibold
(Hrsg.)

Corporate Semantic Web

Wie semantische Anwendungen in
Unternehmen Nutzen stiften

Herausgeber
Börteçin Ege
Technische Universität Wien
Wien
Österreich

Anatol Reibold
OntoPort UG und COGIA GmbH
Darmstadt
Deutschland

Bernhard Humm
Fachbereich Informatik
Hochschule Darmstadt
Darmstadt
Deutschland

ISSN 1439-3107
X.media.press
ISBN 978-3-642-54885-7
DOI 10.1007/978-3-642-54886-4

ISBN 978-3-642-54886-4 (eBook)

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

Springer Vieweg

© Springer-Verlag Berlin Heidelberg 2015

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsgesetz zugelassen ist, bedarf der vorherigen Zustimmung des Verlags. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften. Der Verlag, die Autoren und die Herausgeber gehen davon aus, dass die Angaben und Informationen in diesem Werk zum Zeitpunkt der Veröffentlichung vollständig und korrekt sind. Weder der Verlag noch die Autoren oder die Herausgeber übernehmen, ausdrücklich oder implizit, Gewähr für den Inhalt des Werkes, etwaige Fehler oder Äußerungen.

Gedruckt auf säurefreiem und chlorfrei gebleichtem Papier

Springer Berlin Heidelberg ist Teil der Fachverlagsgruppe Springer Science+Business Media
(www.springer.com)

Vorwort der Herausgeber

„Warum gibt es eigentlich kaum Literatur zu semantischen Anwendungen, die in Unternehmen Nutzen stiften?“ Diese Frage stellten wir uns 2013. Dass es bereits solche Anwendungen gibt, wissen wir aus eigener Projekterfahrung. Aber gerne hätten wir unsere Erfahrungen mit anderen ausgetauscht, die auch bereits erfolgreich semantische Anwendungen für und mit Unternehmen und deren Mitarbeitern und Kunden entwickelt haben.

Als Moderatoren der XING-Arbeitsgruppen „Semantic Web“ (3500 Mitglieder) und „Ontologien in den Informationswissenschaften“ (1250 Mitglieder) streckten wir unsere Fühler aus: das Feedback war enorm. Zum einen gibt es bereits zahlreiche solcher Anwendungen; zum anderen suchen viele nach Hinweisen, Tipps und Tricks zur Entwicklung derselben. Aber trotz der hohen Aufmerksamkeit für Themen rund um das Semantic Web, trotz vieler Artikel und Bücher zu Grundlagen, Standards, Werkzeugen und Visionen, gibt es kaum Literatur, die Architekten, Projektleiter, Ontologie-Modellierer oder Entwickler semantischer Anwendungen bei ihrer Arbeit unterstützt. So war die Idee zum Buch „Corporate Semantic Web – Wie semantische Anwendungen in Unternehmen Nutzen stiften“ geboren.

Corporate Semantic Web – im engeren Sinne bedeutet dies Semantic Web Anwendungen, welche innerhalb eines Unternehmens verwendet werden. Wir möchten den Blick weiten und auch semantische Anwendungen einbeziehen, die in Organisationen aller Art – kommerziell und nicht kommerziell – eingesetzt werden, nicht nur von Mitarbeitern, sondern auch von Kunden oder Partnern. Es geht also um reale Anwendungen für relevante Anwendungsfälle mit echten Anwendern.

Schnell hatten wir über 30 Mitautoren zusammen: Namhafte Experten aus Wissenschaft und Industrie, welche prägende Erfahrungen – positive wie negative – in der Entwicklung semantischer Anwendungen weitergeben. Der Erfahrungshorizont ist groß und umfasst Anwendungen aus so unterschiedlichen Branchen wie Telekommunikation, Logistik, verarbeitende Industrie, Energie, Medizin, Tourismus, Bibliotheks- und Verlagswesen sowie Kultur. Sie berichten über Software-Architektur, Methodik, Projektmanagement, Technologieauswahl, Linked Open Data Sets, Lizenzfragen etc.

Von Anfang an war uns wichtig, dass dieses Buch mehr als eine lose Aufsatzsammlung wird: Ein inhaltlich zusammenhängendes Werk, das Corporate Semantic Web in seinen vielen Facetten beleuchtet. Aus diesem Grund haben wir uns vom 25.-27.6.2014 im renommierten Leibniz-Zentrum für Informatik „Schloss Dagstuhl“ zu einem Workshop zusammengefunden, um alle Kapitel und deren Kernaussagen zu besprechen. Dabei haben wir nicht nur dieses Buch entworfen, sondern sind auch als Community zusammengewachsen und haben viel voneinander gelernt.

Ganz herzlichen Dank geht an unsere geschätzten Mitautoren sowie an Hermann Engesser vom Springer-Verlag, der nicht nur durch seine Teilnahme am Dagstuhl-Workshop gezeigt hat, dass ihm das Thema am Herzen liegt.

Wien und Darmstadt
im November 2014

Börteçin Ege
Bernhard Humm
Anatol Reibold

Inhaltsverzeichnis

1 Corporate Semantic Web	1
Bernhard Humm, Börteçin Ege und Anatol Reibold	
2 Einordnung und Abgrenzung des Corporate Semantic Webs	11
Adrian Paschke und Ralph Schäfermeier	
3 Marktstudie: Welche Standards und Tools werden in Unternehmen eingesetzt?	23
Börteçin Ege	
4 Modellierung des Sprachraums von Unternehmen	41
Thomas Hoppe	
5 Schlendern durch digitale Museen und Bibliotheken	59
Bernhard Humm und Timm Heuss	
6 Semantische Suche im Bereich der Energieforschungsförderung	71
Ulrich Schade, Hermann Bense, Michael Dembach und Lukas Sikorski	
7 Semantische Suche in einer digitalen Bibliothek	87
Natalja Friesen und Christoph Lange	
8 Semantische Beratung im Tourismus-Sektor	101
Bernhard Humm und Oliver Juwig	
9 Messung des Nutzens semantischer Suche	111
Thomas Hoppe und Horst Junghans	

10 Verlage müssen sich neu erfinden	129
Christian Dirschl und Katja Eck	
11 Semantische Technologien für Mobilfunkunternehmen	145
Abdulkaki Uzun und Gökhan Coskun	
12 Semantische Analyse großer Datenbestände aus unternehmensinternen und externen Quellen	167
Pascal Lauria und Thomas Reuter	
13 Intelligente Datenauswertung mit Linked Open Data	187
Heiko Paulheim, Axel Schulz, Frederik Janssen, Petar Ristoski und Immanuel Schweizer	
14 Sicheres Spielzeug für Kinderhände – mit Hilfe semantischer Datenbanken	203
Claudia Baumer, Achim Steinacker und Andreas Weber	
15 Markttopologien mit semantischen Netzen	215
Klaus Reichenberger und Martin Kirsch	
16 Ontologien als Schlüsseltechnologie für die automatische Erzeugung natürlichsprachlicher Texte	227
Hermann Bense und Ulrich Schade	
17 Semantische Technologien und Standards für das mehrsprachige Europa	247
Georg Rehm und Felix Sasaki	
18 Linked Data im digitalen Tanzarchiv der Pina Bausch Foundation	259
Bernhard Thull, Kerstin Diwisch und Vera Marz	
19 Mobile semantische P2P Anwendungen bauen	277
Thomas Schwotzer	
20 Intelligente Wissenswiederverwendung in internationalen Logistik-Projekten	289
Stephan Zelewski, Martin Kowalski und Daniel Bergenrodt	
21 Ontologien für klinische Studien	307
Heinrich Herre und Alexandr Uciteli	

22 Die Rolle der Anfragesprache SPARQL im Kontext von Linked Data . . .	319
Markus Luczak-Rösch	
23 Umbenennungen im Unternehmensalltag	329
Thomas Hoppe	
24 Verteilte und agile Ontologieentwicklung	341
Ralph Schäfermeier und Adrian Paschke	
25 Unterstützung komplexer Entscheidungsprozesse	359
Josef Zelger	
26 Lizenzierung und Nutzung vernetzter Daten – Fallstricke und Empfehlungen	381
Tassilo Pellegrini	
Sachverzeichnis	397

Wie semantische Anwendungen in Unternehmen Nutzen stiften

Bernhard Humm, Börteçin Ege und Anatol Reibold

1. Semantische Anwendungen haben Business Potential und stiften bereits heute in Unternehmen Nutzen, z. B. in den Branchen Telekommunikation, Logistik, verarbeitende Industrie, Energie, Medizin, Tourismus, Bibliotheks- und Verlagswesen und Kultur.
2. Zahlreiche Linked Data Sets stehen zur Verfügung, aber die fachliche Passung zum Anwendungsproblem muss sorgfältig geprüft werden. Manchmal ist die anwendungsspezifische Modellierung empfehlenswert.
3. Semantic Web Technologien sind reif für den Unternehmenseinsatz, aber je nach Anforderungen können auch andere Technologien für die Entwicklung semantischer Anwendungen empfohlen werden.
4. Semantische Anwendungen zu entwickeln erfordert zusätzlich zum klassischen Software Engineering besondere Fähigkeiten, z. B. die eines Knowledge Engineers.

B. Humm (✉)
Hochschule Darmstadt, Darmstadt, Deutschland
E-Mail: Bernhard.Humm@h-da.de

B. Ege
Technische Universität Wien, Wien, Österreich
E-Mail: info@bortecin.com

A. Reibold
OntoPort UG und Cogia GmbH, Darmstadt, Deutschland
E-Mail: anatol.reibold@gmail.com

1.1 Semantic Web

Semantic Web ist seit der Prägung des Begriffs 2001 durch Tim Berners-Lee ein viel benutzter Begriff. Die Vision ist, Daten über Anwendungs-, Firmen- und Ländergrenzen hinweg auszutauschen und wiederzuverwenden. Das World Wide Web Consortium (W3C) hat dazu in den letzten Jahren als Rahmenwerk für das Semantic Web umfangreiche Standards, insbesondere RDF, RDFS, OWL und SPARQL¹, etabliert. Darüber hinaus wurden professionelle Werkzeuge für die Ontologie-Entwicklung sowie für die Speicherung und für den Zugriff auf semantische Daten entwickelt und zur Reife gebracht.

Organisationen weltweit haben zahlreiche so genannte „Linked Open Data Sets“ erstellt und veröffentlicht. Die Grundlagen von Semantic Web sind also etabliert und fundiert beschrieben. Auch mangelt es nicht an visionären Artikeln. Überraschenderweise gibt es jedoch wenig Literatur über semantische Anwendungen, die in Unternehmen bereits eingesetzt werden und Nutzen stiften.

Woran liegt das? Gibt es noch gar keine solchen Anwendungen?

1.2 Semantische Anwendungen im Unternehmenseinsatz

Tatsächlich gibt bereits zahlreiche semantische Anwendungen, die in Unternehmen bzw. Organisationen verschiedenster Branchen im Einsatz sind, z. B. Telekommunikation, Logistik, verarbeitende Industrie, Energie, Medizin, Tourismus, Bibliotheks- und Verlagswesen und Kultur. Von solchen Anwendungen handelt dieses Buch. Abbildung 1.1 gibt eine Übersicht über die Branchen und Themenbereiche der in diesem Buch beschriebenen semantischen Anwendungen.

Die Autoren der einzelnen Kapitel haben umfangreiche Erfahrungen in der Entwicklung von semantischen Anwendungen gemacht – positive wie negative. Sie berichten über Software-Architektur, Methodik, Technologieauswahl, Linked Open Data Sets, Lizenzfragen etc. Manche dieser Erfahrungen und die daraus resultierenden Empfehlungen mögen überraschen – umso wichtiger, dass sie von Praktikern für Praktiker aufgeschrieben sind. Nachfolgend fassen wir wesentliche Erkenntnisse zusammen.

1.3 Bereitstellen von Linked Data reicht nicht

Eine der Kernideen des Semantic Web und von Linked Data ist es, das Rad nicht stets neu zu erfinden, sondern Data Sets zwischen Communities auszutauschen und wiederzuverwenden. Damit das immer besser gelingt, stellt Tim Berners Lee Prinzipien für die Veröffentlichung von Linked Data vor [2]. Diese enthalten die Empfehlungen, HTTP-URIs zur

¹ Für eine praxisnahe Einführung in Semantic Web Technologien siehe z. B. [1].

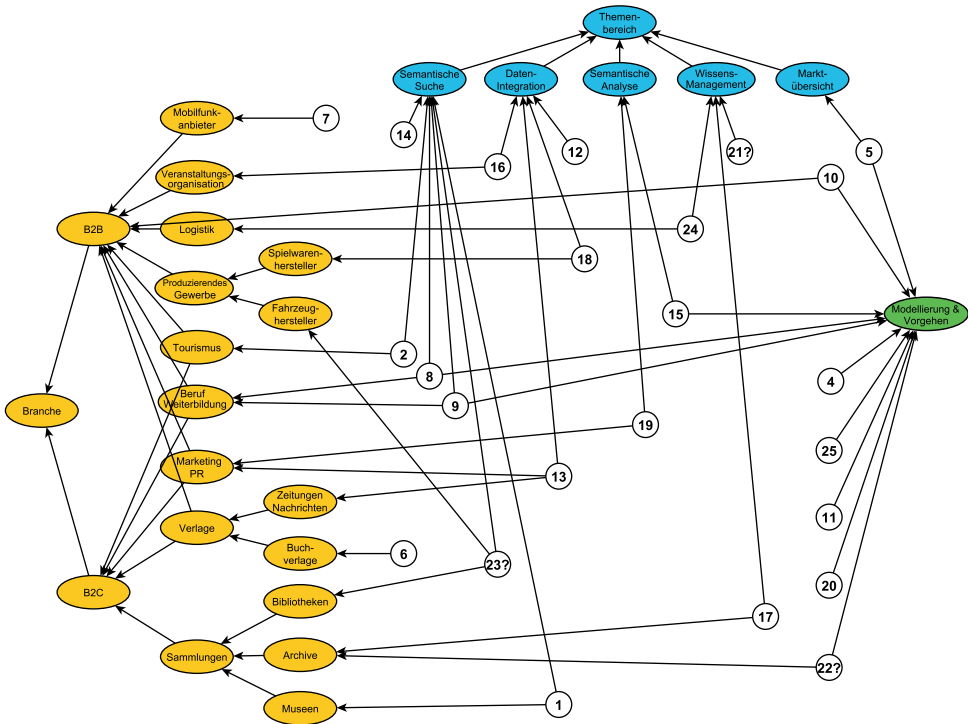


Abb. 1.1 Branchen und Themenbereiche der im Buch beschriebenen semantischen Anwendungen. Die Zahlen sind Kapitelnummern (Graphik von Thomas Hoppe)

Bezeichnung von Objekten zu verwenden, Standards wie RDF und SPARQL zu verwenden sowie zusätzlich zu Objekt-Informationen auch Links auf andere Objekte anzugeben.

Die Festlegung auf Standards ist ein wichtiges Fundament und Design-Prinzipien sind hilfreich, aber reicht das? Wer semantische Anwendungen entwickelt und dabei auf existierende, veröffentlichte Data Sets zugreift, ist mit weiter gehenden Fragestellungen konfrontiert – selbst wenn die o. g. Prinzipien eingehalten wurden.

Da ist zunächst die Frage der *Datenqualität*. Diese kann niemals absolut beantwortet werden, sondern ist immer in Relation zu einem konkreten Nutzungskontext zu sehen. Ein Beispiel: Wie ist die Datenqualität von DBpedia in Bezug auf Künstler und ihre Zuordnung zu kunstgeschichtlichen Epochen? Beispielsweise liefert die Anfrage nach Renaissance-Künstlern nur 19 Resultate². Michelangelo ist vorhanden, Raffael sucht man jedoch vergebens. Statt dessen findet man auch die Resource `dbpedia:Leonardo_da_Vinci's_personal_life`. Es scheint also, dass derzeit die Datenqualität von DBpedia für semantische Anwendungen im Bereich der Kunstgeschichte nicht ausreichend ist.

² SPARQL Query `select ?p where {?p rdf:type yago:RenaissanceArtists}` ausgeführt auf dem SPARQL Endpoint `http://dbpedia.org/sparql` am 15.7.2014.

Die zweite Frage bezieht sich auf den *Modellierungskontext*. Bowker and Star schreiben ganz richtig in [3], dass Klassifikationen, die in einem Kontext wunderbar natürlich und homogen erscheinen, in einem anderen Kontext gezwungen und uneinheitlich wirken. Es gibt halt nicht „die richtige“ Modellierung eines Sachgebiets – es hängt vom Nutzungskontext ab.

Bei der Integration mehrerer Data Sets stellt sich die Frage der *Interoperabilität*. Dies ist eine Frage der Modellierung, die sich auch stellt, wenn die oben genannten Prinzipien strikt eingehalten wurden. Betrachten wir beispielsweise zwei Thesauri, beide in RDFS modelliert, aber von unterschiedlichen Teams nach unterschiedlichen Modellierungsvorschriften. Im einen Thesaurus sind Begriffe als Instanzen modelliert und die Beziehung zwischen Ober- und Unterbegriffen mittels SKOS [4] (skos:broader bzw. skos:narrower); im anderen Thesaurus sind Begriffe als Klassen modelliert und dieselbe semantische Beziehung über eine andere Property, nämlich rdfs:subclassOf. Werden beide Thesauri integriert, so sind einfache SPARQL-Abfragen über den integrierten Thesaurus nicht mehr möglich. Weitaus schwerwiegender als diese technischen Interoperabilitätsprobleme sind jedoch die fachlichen: Fragen der Redundanz, Konsistenz, Kohärenz, des Mappings und der unterschiedlichen Vollständigkeitsgrade der integrierten Thesauri.

Mit diesen Fragestellungen beschäftigen sich u. A. Kapitel 5, 8 und 13.

1.4 Eine global vernetzte Wissensbasis – Fiktion oder Realität?

Zu vielen semantischen Data Sets werden SPARQL Endpoints zur Abfrage bereit gestellt. Beispiele sind DBpedia, DBLP und Gene Ontology Database. Außerdem erlaubt der SPARQL-Standard den Zugriff auf verschiedene Endpoints und die Integration der Ergebnisse. Die Idee ist daher naheliegend, in semantischen Anwendungen auf die lokale Datenhaltung zu verzichten und statt dessen die Daten über verteilte SPARQL Queries zu integrieren – ganz im Sinne einer service-orientierten Architektur. Das spart lokalen Speicherplatz, erübrigt die Installation eines RDF Triple Stores und garantiert einen stets aktuellen Datenbestand.

In semantischen Anwendungen, die im Praxiseinsatz sind, wird diese Form der Integration jedoch selten gewählt. In diesem Buch werden 18 semantische Anwendungen vorgestellt, die in Unternehmen eingesetzt werden. Lediglich zwei davon greifen über SPARQL Endpoints auf verteilte Data Sets zu.

Wie kommt das? Zum Einen erschweren die oben erwähnten Problemfelder Datenqualität, Modellierungskontext und Interoperabilität eine einfache Datenintegration über SPARQL-Endpoints. Dazu kommen Fragen der Performanz und Verfügbarkeit. Hier gilt, dass die Kette nur so stark wie ihr schwächstes Glied ist. Ist nur einer von vielen Endpoints nicht verfügbar, so steht u. U. die ganze Anwendung. Und Performanz aktueller SPARQL-Endpoints ist meist nicht ausreichend für Online-Anwendungen, in denen Endnutzer-Antwortzeiten von höchstens einer Sekunde gefordert sind.

Kapitel 22 behandelt detailliert Vor- und Nachteile von SPARQL-Endpoints.

1.5 Semantik=RDF?

Eine wesentliche Voraussetzung für den Erfolg der Idee des Semantic Web ist die konsequente Standardisierung von Sprachen wie RDF, RDFS, OWL und SPARQL, die wesentlich vom W3C vorangetrieben wurde. Sie erlaubt erst die verteilte, unabhängige Entwicklung von integrierbaren Data Sets im Rahmen eines Community Prozesses. Sie erlaubt auch erst die Entwicklung von Werkzeugen, z. B. Ontologie-Editoren und RDF Triple Stores, deren Daten austauschbar sind.

Viele semantische Anwendungen verwenden direkt diese Standards und Werkzeuge. Von den in diesem Buch vorgestellten 18 Anwendungen verwenden 8 RDF Triple Stores. Das bedeutet aber auch im Umkehrschluss, dass in etwa der Hälfte der Anwendungen bewusst auf andere Technologien gesetzt wird. Beispiele für die Nutzung anderer Technologien finden sich bei der semantischen Suche für Bibliotheken und Museen (Kap. 5), Analyse von Marktdaten (Kap. 12) und einem Semantic Social Network für Sport-Trainer (Kap. 19).

Sind das damit keine semantischen Anwendungen? Wird eine Anwendung semantisch, wenn ein Semantic Web Standard des W3C verwendet wird, und verliert sie diese Eigenschaft, wenn sie dies nicht der Fall ist? Aus unserer Sicht gilt dies keinesfalls. Eine Anwendung kann als semantisch bezeichnet werden, wenn die *Bedeutung* von Inhalten eine wesentliche Rolle spielt. Das betrifft den Anwender und ist unabhängig von der eingesetzten Technologie.

Warum entscheiden sich Architekten solcher semantischer Anwendungen bewusst, bei der Entwicklung auf andere Technologien zu setzen? Dafür gibt es mehrere Gründe. Zum einen können Performanz-Gründe eine Rolle spielen, wenn es sich um sehr große Datenmengen und gleichzeitig hohen Anforderungen an die Antwortzeiten handelt (siehe Kap. 5). Weiterhin kann die Funktionalität von RDF Triple Stores für den Anwendungsfall nicht ausreichend sein. Beispiel sind die Möglichkeiten für Volltextsuche, Teilwortsuche, phonetische Suche oder Toleranz gegenüber Tippfehlern. Die RDF-Technologie kann auch das Know-How des Entwicklerteams übersteigen und der Know-How-Aufbau kann als zu kostspielig eingeschätzt werden. In diesem Fall können einfachere Technologien eingesetzt werden, wenn sie dem Anwendungsfall angemessen sind (siehe Kap. 19). Sich gegen Semantic Web Technologien zu entscheiden kann aber auch Nachteile haben, z. B. ein erschwertes Austausch von Daten oder die Bindung an eine proprietäre Technologie.

1.6 Richtig vorgehen

Wie in jedem Entwicklungsprojekt steht am Anfang der Entwicklung einer semantischen Anwendung die Kundenerwartung. Wer sind die Anwender? Sind sie unternehmensintern oder Endkunden? Was sind die Anwendungsfälle? Und welche Fragen soll die Anwendung beantworten können. Es empfiehlt sich, solche Kompetenzfragen konkret zu sam-

meln und aufzuschreiben, z. B. „Wie verteilen sich die Fördergelder für Solarkraftwerke auf die Bundesländer?“ (Kap. 6 und 20).

Gerade für semantische Anwendungen gilt, dass sich zukünftige Anwender vorab gar nicht genau vorstellen können, wie die Anwendung funktionieren soll. Daher ist meist ein agiles Vorgehen empfehlenswert: klein anfangen, rasch Prototypen entwickeln, intensiv die Anwender einbeziehen und dann inkrementell ausbauen (siehe Kap. 10 und 24).

In Ergänzung zu den Rollen in klassischen Software-Entwicklungsprojekten wird die Rolle des Knowledge Engineers wichtig. Er ist verantwortlich für die Akquisition und Formalisierung des Wissens in Zusammenarbeit mit den Fachexperten. Meist geschieht dies in Form von Interviews. Dabei empfiehlt es sich, die Fachexperten in Form von „narrativen Stories“ frei über ihre Domäne berichten zu lassen und diese Berichte zu dokumentieren (siehe Kap. 20). Die Arbeit am Modellierungswerkzeug übernimmt dann meist der Knowledge Engineer selbst. Die Erfahrung zeigt, dass Fachexperten selten mit Modellierungswerkzeugen umgehen können. Häufig verstehen sie nicht die Notation, Konzepte und Vorgehensweisen und zeigen auch keine Bereitschaft, sich dieses Wissen anzueignen. Falls Fachexperten doch Werkzeuge benutzen sollen, dann sollten diese Werkzeuge in der Bedienung so einfach wie möglich sein, z. B. ein Spreadsheet, welches in der täglichen Arbeit ohnehin vertraut ist (siehe Kap. 15).

Wichtig ist auch die Qualitätssicherung durch Zurückspielen des entstehenden Modells an die Fachexperten. Wie kann aber Qualitätssicherung des Modells funktionieren, wenn die Fachexperten die Modellnotation nicht verstehen? Hier haben sich Visualisierungswerkzeuge bewährt, wie z. B. hyperbolische Bäume (siehe Kap. 6).

1.7 Modellieren ist einfach (!)

Make or buy? Diese Frage stellt sich bei der Entwicklung semantischer Anwendungen nicht nur für technische Produkte und Werkzeuge, sondern auch für Data Sets. Linked Open Data Sets beinhalten hundertausende von Begriffen, in jahre- oder jahrzentelanger Arbeit von Experten formalisiert und qualitätsgesichert. Diesen Schatz gilt es zu heben.

Aber dennoch zeigt es sich in der Praxis immer wieder, dass keines der öffentlichen Data Sets genau auf das aktuelle Problem passt. Und dann muss man sorgfältig die Alternativen abwägen, a) mit den Schwächen zu leben, b) verschiedene öffentliche Data Sets zu kombinieren, c) diese anzureichern – evt. im Rahmen eines Community Prozesses – oder d) doch ein anwendungsspezifisches Data Set neu zu entwickeln. In der jetzigen relativ frühen Phase der Technologie hat sich auch bewährt, mit dem Hersteller eines Data Set in Kontakt zu treten, um evtl. Anpassungen direkt zu besprechen.

Die Entwicklung eines Data Sets fordert das Know-How des Knowledge Engineers. Wichtig ist, dass die Modellierung nicht im luftleeren Raum, sondern stets im Kontext einer konkreten Fragestellung erfolgt. Beispielsweise muss eine Ontologie für die semantische Suche von Hotels Konzepte wie Ausstattungsmerkmale, Sehenswürdigkeiten etc. umfassen. Sie braucht aber keine Aussagen zu beinhalten wie „Im Hotel arbeiten Men-

schen“, „Menschen (Homo Sapiens) gehören zur Ordnung der Primaten, zur Klasse der Säugetiere, etc., haben ein Herz, eine Lunge, eine Leber etc. etc.“ Diese Aussagen – obwohl wahr – in die Ontologie aufzunehmen, wäre nicht nur unnötig, sondern im Sinne der Anwendung sogar falsch. Sie nutzen nicht dem Anwendungszweck und verursachen bei der Entwicklung und Pflege nur unnötige Kosten.

Wichtig ist also nicht nur, zu wissen, was noch fehlt, sondern auch zu wissen, wann man aufhören kann. Ein Plädoyer für „Mut zur Lücke“ findet sich in Kap. 4. Werden solche Hinweise beachtet, dann muss Modellieren gar nicht so aufwändig sein, wie gemeinhin vermutet. Mit einem eingespielten Team ist die Modellierung von 4000 Begriffen in ca. 20 Arbeitertagen realistisch.

1.8 Juristische Fragen

Sollen Linked Data Sets in eine semantische Anwendung eingebunden werden, dann stellt sich die Frage, ob die entsprechenden Lizenzen dies auch erlauben. Vernetzte Datensätze und Werke, die auf diesen Datensätzen aufbauen, sind nach dem Immaterialgüterrecht geschützt. Die Datensätze, aus denen sich Linked Data speist, bedienen sich entsprechend ihrer Offenlegungspolitik einer Vielzahl an Lizenzmodellen mit teils sehr unterschiedlichen Terminologien, Geltungs- und Gültigkeitsbereichen. Bei der Wiederverwendung „offener“ Daten ist auf die Kompatibilität der Lizenzbedingungen der Datensätze zu achten. Lizenzkonflikte können die Wiederverwendung einschränken und zu juristischen Folgeproblemen führen. Darauf geht Kap. 26 ein.

1.9 Semantische Anwendungen stiften Nutzen in Unternehmen – nachweislich!

Unternehmen werden nach betriebswirtschaftlichen Gesichtspunkten geführt. Nur was sich lohnt, wird auch gemacht. Semantische Technologien sind noch relativ neu und Know-How-Aufbau kostet Zeit und Geld. Lohnt sich diese Investition? Steht den Kosten ein ausreichender Nutzen semantischer Anwendungen gegenüber? Kann dieser Nutzen gemessen und nachgewiesen werden? Damit beschäftigt sich Kap. 9.

Die Praxis zeigt, dass eine seriöse Quantifizierung des Nutzens semantischer Anwendungen – hier: semantische Suche – gar nicht so einfach ist. Klassische Maße wie Precision und Recall aus dem Information Retrieval, die auf einem Goldstandard basieren, sind nicht geeignet. Semantische Suche liefert einerseits weniger, dafür aber genauere Treffer; andererseits aber auch zusätzliche Treffer. Die Bewertung einer semantischen Suche muss diesen vermeintlichen Widerspruch auflösen. Daher wird ein neues Maß für die quantitative Bewertung der Effizienzsteigerung bei der Suche empfohlen. Empirische Vergleiche zeigen, dass durch semantische Suche Effizienzsteigerungen seitens des Benutzers in der Größenordnung von 10–15 % möglich sind.

1.10 Fazit

Semantische Anwendungen stiften schon heute in Unternehmen Nutzen. Die zugrunde liegenden Technologien haben die dafür notwendige Reife erreicht. Auch sind in den letzten Jahren umfangreiche Linked Data Sets entstanden. Die Entwicklung semantischer Anwendungen erfordert aber Know-How in Ergänzung zum klassischen Software Engineering Know-How.

In den folgenden Kapiteln geben Experten ihre Erfahrungen in der Entwicklung semantischer Anwendungen weiter. Kapitel 1–4 geben einen Überblick über das Feld, jeweils aus unterschiedlichen Gesichtspunkten. Kapitel 5–21 beschreiben semantische Anwendungen und was man daraus lernen kann. Kapitel 22–26 geben Hinweise zu methodischen und querschnittlichen Fragen.

Literatur

1. Dean, Allemang, und James Hendler. 2011. *Semantic web for the working ontologist-effective modeling in RDFS and OWL*. 2nd ed. San Francisco: Morgan Kaufmann Publishers Inc. (978-0-12-385965-385965).
2. Tim, Berners-Lee. Linked data. <http://www.w3.org/DesignIssues/LinkedData.html>. Zugegriffen: 15. Juli 2014.
3. Bowker, G. C., und S. L. Star. 1999. *Sorting things out: classification and its consequences*. Cambridge: MIT Press.
4. SKOS (Simple Knowledge Organization System). www.w3.org/2004/02/skos/. Zugegriffen: 15. Juli 2014.



Bernhard Humm ist Professor für Software Engineering und Projektmanagement am Fachbereich Informatik der Hochschule Darmstadt und Direktor des Instituts für Angewandte Informatik Darmstadt (aIDa). Seine Forschungsschwerpunkte liegen in den Bereichen Semantic Web, Software Architektur und Programmiersprachen. In diesen Bereichen führt er Forschungsprojekte im Industrie- und Hochschulumfeld durch und publiziert regelmäßig. Prof. Humm arbeitete 11 Jahre beim Software- und Beratungshaus sd&m AG als technischer Senior- und Chefberater, Bereichsleiter und Leiter von sd&m Research. Seine Promotion erhielt er an der University of Wollongong in Australien, sein Diplom mit Auszeichnung an der Universität Kaiserslautern.



Dipl.-Ing. Börtçin Ege studierte Informatik an der Technischen Universität Wien. Er beschäftigt sich seit 2005 mit Semantic Web. In 2007 gründete er die Arbeitsgruppe Semantic Web in XING und ist seit mehreren Jahren in Semantic Web Projekten aktiv tätig. Er ist Doktorand an der TU Wien und auch Autor von zahlreichen wissenschaftlichen Publikationen. Seine Forschungsschwerpunkte sind Corporate Semantic Web, Future Internet und Industrie 4.0. Mehr über seine Arbeiten finden Sie unter <http://www.bortecin.com/>.



Anatol Reibold studierte Mathematik und Mechanik an der Universität Novosibirsk. Leidenschaftlicher Mathematiker mit Expertise in Ontologien, Semantic Web, Datenanalyse, Data Science, Informationswissenschaften, Kryptographie, GABEK, Polykontexturale Logik, Polyduktive Theorie. Zusammen mit Freunden gründete er den Darmstädter Ontologenkreis sowie die XING-Gruppen „Ontologien in den Informationswissenschaften“ und „Denkkunst Schule“.

Adrian Paschke und Ralph Schäfermeier

1. **Semantic Web** adressiert die Repräsentation und die Bedeutung von Informationen und Wissen im Web
2. **Corporate Semantic Web (CSW)** wendet semantische Technologien im Unternehmenskontext an, sowohl zur Unterstützung interner (Web-basierter) Geschäftsinformationssysteme und –prozesse, als auch in Anwendungen und Diensten, die im öffentlichen Web (**Public Semantic Web**) von Unternehmen angeboten werden
3. Das **Social Semantic Web** oder auch **Web 3.0**, als Verbindung von kollaborativen Web 2.0 mit Semantic Web, unterstützt sozial Interaktionen semantisch und findet insbesondere in der Kollaboration in Unternehmen seine Anwendung im Corporate Semantic Web
4. Das **Pragmatic Web** überlagert die reine Semantik, im Sinne der Bedeutung von Wissen, mit pragmatischen Verständnis und Verpflichtungen, wie sie z. B. in einem Interaktionskontext entstehen, d. h. es geht auch um die Fragen „warum“ und „wie“ werden Informationen und Wissen (in einem bestimmten pragmatischen Kontext) genutzt.
5. Neben technologischen Aspekten müssen im Corporate Semantic Web auch der pragmatische Kontext der Entwicklung und Nutzung semantischen Wissens im Unternehmen berücksichtigt werden, zum Beispiel ökonomischer Kontext wie Kostenmodelle für die Ontologieentwicklung (siehe Kap. 24) oder Rechte und Pflichten entsprechen der Rollen in Projekten, Prozessen, Organisationsstruktur, etc.

A. Paschke (✉) · R. Schäfermeier
AG Corporate Semantic Web, Freie Universität Berlin, Berlin, Deutschland
E-Mail: paschke@inf.fu-berlin.de

6. Ein möglicher „evolutionärer“ nächster Schritt des Webs kann die Entwicklung hin zu einem aktiven **Pragmatic Web 4.0** sein, in dem z. B., Inhalte nicht länger passiv sind, sondern sich aktiv dem pragmatischen Interaktionskontext anpassen (Stichwort im CSW: **Corporate Smart Content**). Dies würde über die reine Wissensrepräsentation im Semantic Web hinaus eine Weiterentwicklung der Maschinenintelligenz, sowie das Erlernen zugrunde liegender pragmatischer Regeln im Sinn von Weisheit, erfordern.

Das Corporate Semantic Web fokussiert auf die Anwendung semantischer Web Technologien und Methoden im Unternehmensumfeld [1–4]. In diesem Kapitel sollen die Beziehungen des Corporate Semantic Web zu angrenzenden Anwendungsgebieten geklärt werden. Dazu wird zunächst eine grundlegende Begriffsklärung vorgenommen, um darauf aufbauend das Corporate Semantic Web gegenüber anderen verwandten Gebieten, wie dem Public Semantic Web, Social Semantic Web und Pragmatic Web, einzuordnen.

2.1 Grundlegende Begriffe

Von zentraler Bedeutung in Unternehmen sind Informationen und Wissen, welche für Entscheidungen und das Management auf operationaler, taktischer und strategischer Ebene benötigt werden. Bei **Informationen** handelt es sich um **Daten**, die für einen bestimmten (Entscheidungs-)kontext aufbereitet und in Zusammenhang gebracht wurden und so eine Bedeutung und eine Nützlichkeit für einen Informationsnutzer erlangen. Die Interpretation der Informationen erfordert dabei **Wissen**, das entweder implizit oder explizit vorliegen kann. So handelt es sich beispielweise bei den Summen 1 Mio EUR, 1,5 Mio EUR, 2,0 Mio EUR um Daten. Werden diese in Zusammenhang mit den Jahren 2012, 2013, 2014 als Unternehmensgewinn gebracht, stellen diese Informationen über den Unternehmenserfolg dar. Eine Interpretation ist nun mit weiterem Wissen, zum Beispiel über die Entwicklung des Marktes und entsprechender Wettbewerber, möglich, z. B. als dynamischer Gewinnwachstum des Unternehmens. Werden dabei tiefere Zusammenhänge verstanden und können daraus grundlegende Prinzipien abgeleitet werden, wie z. B. Wirkungsprinzipien, spricht man von **Intelligenz** oder auch **Weisheit** (Schlagwort: „Smart ...“).

Daten-, Informations- und Wissensrepräsentation beschäftigt sich mit deren maschineller Abbildung, Verarbeitung und Verwaltung. Während **unstrukturierte** Repräsentationsformate, wie z. B. natürlich-sprachlicher Text oder grafische Darstellungen, die Nutzung durch Menschen erleichtern, werden für die effiziente maschinelle Verarbeitung **strukturierte** oder **semi-strukturierte** Formate benötigt. Derartige Formate beschreiben dabei häufig neben der reinen syntaktischen Struktur (**Syntax**) auch explizit deren **Semantik**, so dass Maschinen eine automatische Bedeutungsinterpretation möglich ist. Erlernen Maschinen dabei durch künstliche Intelligenz neben der reinen semantischen Bedeutung auch, wie diese in einem bestimmten Handlungs- oder Interaktionskontext zu verstehen

und anzuwenden ist, spricht man von **Pragmatik**. So kann beispielsweise ein automatisches Business Activity Monitoring (BAM) System wichtige Aktivitätsinformationen als eine Ampelschaltung repräsentieren – syntaktisch mit den drei Farben „grün“, „gelb“ „rot“, semantisch mit der Interpretation „grün=unkritisch“, „gelb=kritisch“, „rot=Störung“. Die Pragmatik kann diese semantische Bedeutung zusätzlich mit einem tieferen Verständnis und kontextuellen Handlungen versehen, so z. B., dass trotz einer Störung zwischen Mitternacht und vier Uhr keine sofortige Reaktion nötig ist, da hier typischerweise Wartungsarbeiten am System vorgenommen werden.

In Abb. 2.1 stellt diesen Zusammenhang bei der Transformation von Daten zu Informationen, Wissen und Intelligenz in Beziehung zur Syntax, Semantik und Pragmatik der Repräsentation dar [5]. Mit zunehmenden maschinellen Verständnis erhöhen sich gleichzeitig auch die Dichte der Wissensverknüpfungen und die Beständigkeit der zu Grunde liegenden semantischen und pragmatischen Regeln. Während mit, auf formaler Logik basierenden, Ontologie- und Regelsprachen bereits Repräsentationsformalismen existieren, die ein automatisches Maschinenverständnis erlauben, ist die Frage nach dem Erlernen und Verstehen zu Grunde liegender Pragmatik als Basis für echte künstliche Intelligenz und Weisheit ein offenes Forschungsgebiet, das über die klassische formale Logik hinausgeht.

Heutige Informationssysteme sind meist Web-basiert. Die Repräsentationssprachen im Web können in unstrukturierte (z. B. HTML), semi-strukturierte (z. B. XML) und strukturierte (z. B. Semantic Web mit RDF, RuleML/RIF, RDFS/OWL) Formate unterschieden werden. Während erstere, wie HTML, auf die Informationsrepräsentation für den Menschen abzielen, ist XML ein industriell weithin akzeptierter Standard für den maschinellen Daten- und Informationsaustausch, welcher die maschinelle Verarbeitung der XML



Abb. 2.1 Zusammenhang „Daten-Information-Wissen“

Syntax durch die explizite Schemainformationen ermöglicht. Semantic Web Sprachen können XML als Syntax nutzen und versehen diese auf Basis formaler Logiken um eine für Maschinen auswertbare Semantik mit dem Ziel der expliziten Wissensrepräsentation. Standardisiert im W3C Semantic Web Stack werden dafür insbesondere folgende Wissensrepräsentationstechnologien¹:

- **Web Daten:** Web Daten werden als verteilte Ressourcen im Web identifiziert und können über Verlinkung (Linked Data) in Relation zu anderen Ressourcen und Datenwerten gebracht werden. Die Semantik wird hierbei häufig in der Form von Metadaten erfasst, d. h. mittels Metadatenvokabulare mit bekannter Semantik.
Sprachen: Resource Description Framework (RDF); Simple Protocol and RDF Query Language (SPARQL)
- **Web Regeln:** Web Regeln leiten aus vorhandenem Wissen (Fakten, Daten) neues Wissen für Entscheidungen (Entscheidungsregeln) oder Handlungen (Reaktionsregeln) ab.
Sprachen: Rule Markup Languages (RuleML: RuleML 1.0, Semantic Web Rule Language (SWRL), Rule Interchange Format (RIF))
- **Web Ontologien:** Ontologien beschreiben die Konzepte einer Domäne und deren Relationen untereinander. Sie haben somit das Ziel, Domänenwissen zu repräsentieren, z. B. um damit Typisierungen und Konzeptbeziehungen abzuleiten.
Sprachen: RDF Schema (RDFS), Web Ontology Language (OWL)

Zusammenfassend kann also auch im Web zwischen unterschiedlichen Repräsentationssprachen unterschieden werden, die in Web-basierten Informationssystemen und Wissensmanagementlösungen in Unternehmen genutzt werden können.

2.2 Corporate Semantic Web

Das Corporate Semantic Web baut auf den existierenden Daten-, Informations- und Wissenssystemen im Unternehmen auf [1–4]. Es unterstützt und erweitert diese, sowohl für die Erzeugung von relevanten Unternehmensinhalten (**Semantic Content**) und semantischen Wissensbasen (**Semantic Knowledge**), als auch für die Anwendungen und Verbraucher (**Semantic Applications**) die diese Informationen und das semantische Wissen nutzen wollen. Dabei kann es sich sowohl um Mitarbeiter als auch automatisierte Dienste z. B. in Geschäftsprozessen handeln.

Die Nutzung semantischer Technologien lässt sich dabei in drei Anwendungssäulen unterteilen: Semantische Entwicklung (**Corporate Semantic Engineering**), semantische

¹ Neben den W3C Semantic Web existieren weitere Standardisierungen, wie z. B. OMG SBVR, OMG PRR, OMG API4KB und OntoIOP, OMG ODM, RuleML, OASIS LegalRuleML, ISO CL,

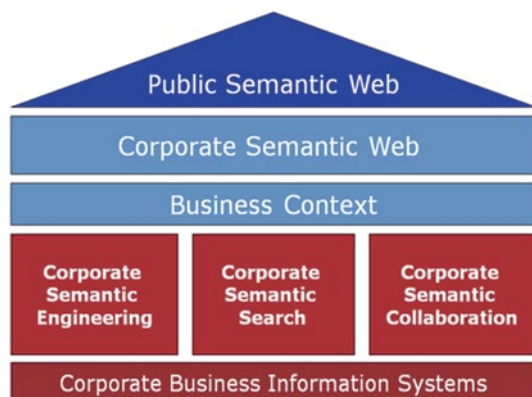
Suche (**Corporate Semantic Search**), semantische Kollaboration (**Corporate Semantic Collaboration**):

- **Corporate Semantic Engineering**
 - Methoden und Werkzeuge zur Entwicklung semantischer Wissensbasen in Form von Ontologien und Regeln
 - Nutzung semantischer Wissensbasen als Hintergrundwissen in Entwicklungs- und Erzeugungsprozessen wie der Erzeugung von Informationen aus Daten
- **Semantic Corporate Search**
 - Lösungen zur semantischen Suche in kontrollierten Daten- und Informationsbeständen mit definierter Qualität und Dienstverbesserungen
- **Semantic Corporate Collaboration**
 - Semantische Kollaborations- und Unterstützungsplattformen für die Zusammenarbeit unterschiedlicher Unternehmensdomänen und Teams

Ein Corporate Semantic Web adressiert aber nicht nur die technologischen, sondern auch die **pragmatischen Aspekte** der Anwendung im Unternehmen. Methoden und Werkzeuge müssen so bereitgestellt werden, dass diese im Unternehmen gut angenommen und leicht in die vorhandene Prozess-, IT-Infra- und Dienststruktur integriert werden können. Von zentraler Bedeutung ist dabei auch der unternehmerische Kontext, z. B. im Sinne ökonomischer Kostenmodelle bei der Entwicklung semantischer Wissensbasen (siehe Kap. 24).

Abbildung 2.2 zeigt diesen skizzierten Aufbau eines Corporate Semantic Web. Entsprechend den Anforderungen moderner Web-basierter Unternehmensanwendungen interagiert ein Corporate Semantic Web dabei auch auf vielfältige Weise mit dem Public Semantic Web, z. B. zur Anreicherung der internen Unternehmensdaten mit öffentlich

Abb. 2.2 Schematischer Aufbau eines Corporate Semantic Web. (Quelle: [4])



vorhandenen Daten und Wissen (wie z. B. Wikipedia/DBPedia Deutsch² aus der Linked Open Data Cloud).

2.3 Public Semantic Web

Das sogenannte öffentliche Semantic Web (**Public Semantic Web**) umfasst alle öffentlich zugänglichen Semantic Web Wissensbasen und Anwendungen. Während „Corporate Semantic Web“-Daten häufig nicht öffentlich oder nur über spezielle Schnittstellen (sog. **Deep Web** Schnittstellen) außerhalb der Unternehmensgrenzen zugänglich sind, liegen die Daten und das Wissen im Public Semantic Web (oder auch Surface Web) öffentlich vor³. Hieraus ergeben sie folgende wesentlichen Unterschiede zwischen dem Public Semantic Web und einem Corporate Semantic Web:

- **Geschlossene** Informationssysteme und Intranetlösungen mit im Unternehmen bekannten und dokumentierten Schnittstellen zwischen den Systemen, Diensten und Domänen.
- **Bekannte** Nutzergruppen innerhalb der Unternehmensnetzwerke.
- Nutzung der **existierenden** Unternehmensinfrastrukturen, Daten, Informationen und Wissen beschränkt innerhalb eines existierenden pragmatischen Kontexts, wie Geschäftsregeln, -politiken und -prozesse.
- Datensicht: Geschlossen, oft strukturiert oder zumindest mit Metadaten angereichert, **bekannte Datenmodelle** (z. B. relationale Datenbankschema, Objekt-orientiert Modelle, XML Schemas etc.).
- Logische Sicht: Partielle Annahme einer **geschlossenen Weltsicht** (d. h. Daten werden als vollständig angenommen und nicht vorhandene Daten als falsch), Annahme **eindeutiger Namensbezeichner** (z. B. Datenbankschlüssel), Erzeugung von Informationssichten auf die Daten (z. B. Data Warehouses).

Diese tendenziellen Unterschiede eines Corporate Semantic Webs erleichtern den Einsatz semantischer Technologien im Unternehmen und lassen sich gezielt für effizientere und passgenauere semantische Erweiterungen der existierenden und bekannten IT-System im Unternehmen nutzen. Selbstverständlich sind hier die Übergänge zum Public Semantic Web fließend und mit der Größe eines Corporate Semantic Webs treten ähnliche Herausforderungen wie im Public Semantic Web auf. Umgekehrt erlaubt ein Corporate Semantic Web Unternehmen seine öffentlichen Daten, Informationen und Wissen mit expliziten semantischen Hintergrundwissen anzureichern und so im Public Semantic Web automatisiert nutzbar zu machen.

² <http://de.dbpedia.org>.

³ Was nicht heißt, dass diese frei verfügbar sein müssen, sondern sie können durch Nutzungslizenzen geschützt sein.

2.4 Social Semantic Web 3.0

Während das Semantic Web Wissen⁴ und damit Maschinen miteinander verknüpft, unterstützt das soziale Web (**Social Web** oder auch **Web 2.0** [9]) Menschen in ihren sozialen Kommunikationen und Relationen untereinander. Die Erweiterung um semantische Webtechnologien wird dann als **Social Semantic Web 3.0** [10] bezeichnet, in dem Menschen, Dienste und Dinge miteinander interagieren und kommunizieren, d. h. ein Web „kollektiver Wissenssysteme“. Im Fokus steht hier somit die soziale Interaktion, unterstützt durch Strategien, Methoden und Technologien des Semantic Webs. Die dabei eingesetzten „Web 3.0“-Werkzeuge ermöglichen insbesondere auch die Entwicklung explizierter semantischer Wissensrepräsentationen, die ausgetauscht und als Basis für ein gemeinsames Verständnis genutzt werden. Nachfolgende Tab. 2.1 zeigt dies exemplarisch anhand dreier typischer Funktionalitäten des Social Webs, den *Tagging*, *Mesh-Ups* und der *Suche*.

Das Social Semantic Web findet sowohl im Public Semantic Web als auch im Corporate Semantic Web seine Anwendung. Es ist vom Corporate Semantic Web nicht abzugrenzen, sondern fokussiert auf die zu anfangs eingeführten Säulen der semantischen Kollaboration (Semantic Collaboration). Detaillierter ist hier zu unterscheiden zwischen:

- **Kommunikation**
 - **Wie** kommunizieren Teilnehmer, IT-Dienste oder IT-Agenten/Dinge miteinander?
- **Kollaboration**
 - **Warum** kollaborieren sie zur Erreichung gemeinsamer Ziele?
- **Koordination**
 - **Was** wird koordiniert im Daten-/Informations- und Prozessfluss?

Tab. 2.1 Web 2.0 im Vergleich zum Web 3.0. (Quelle: Social-Semantic Web [10])

	Web 2.0 (social web)	Web 3.0 (social semantic web)
Tagging	Annotation mit <i>mehrdeutigen</i> Schlüsselwörtern Sprachprobleme: Singular/Plural, Wortstamm, Synonyme, etc. Keine Intelligenz	Annotation mit semantisch <i>eindeutigen</i> Entitäten (z. B. aus Ontologien) <i>Automatische Inferenz</i> (z. B. Subklassen, Synonyme, ...)
Mesh-ups	Manuelle Rekombination von Daten unterschiedlicher Quellen durch „Mesh-ups“ im Voraus (d. h. <i>statisch</i>)	<i>Dynamische</i> Zusammenführung durch Endnutzer und Dienste (<i>Realzeit</i>)
Suche	Schlüsselwort- oder Tag-basierte Suche (<i>syntaktische Suche</i>)	<i>Semantische Suche</i> mit semantischen Ähnlichkeiten und semantischen Vorschlägen

⁴ Stichworte: Metadaten, Linked (Open) Data, ontologische Wissensbasen, verteilte Regelsystems und Agenten.

Werkzeuge des Social Semantic Webs, die im Corporate Semantic Web zum Einsatz kommen, sind typischerweise Erweiterungen vorhandener kollaborativer Wissensmanagementlösungen, wie Groupware Systeme, Social Software (Wikis, Weblogs, Social Bookmarking), Document/Content Management Systeme (kollaborative CMS und DMS, Wissensportale), kollaborative Workflow und Prozesssysteme, etc.

Neben der Semantik steht bei sozialen Interaktionen die Pragmatik im Vordergrund.

2.5 Pragmatic Web

Während das syntaktische Web 1.0 und das semantische Web 2.0 und 3.0 auf die Repräsentation und die Wissensbedeutung von Informationen abzielt, befasst sich das pragmatische Web (**Pragmatic Web**) mit den Werkzeugen, Praktiken und Theorien, die beschreiben *warum* und *wie* Informationen und Wissen in Webinteraktionen (Mensch-Mensch, Mensch-Maschine, Maschine-Maschine) genutzt werden [5, 6]. Dieser **pragmatische Interaktionskontext** führt zu einem situativen **Verständnis** und kontextuellen **Verpflichtungen**. So wird beispielsweise in einer Systeminteraktion auf eine Frage eine Antwort erwartet, selbst dann, wenn die Antwort dem System nicht bekannt sein sollte. Vielfältige Verpflichtungen ergeben sich auch aus dem jeweiligen Kontext, in dem die Interaktion stattfindet, wie z. B. dem Organisationskontext (Organizational Semiotics), der Rolle und deren Rechte und Pflichten, dem Prozesskontext, dem exogenen Geschäftskontext, juristischen Rahmenbedingungen, etc.

Wie das Semantic Web eine Erweiterung des Syntactic Web ist, ist das Pragmatic somit eine Erweiterung des Semantic Webs (insbesondere des Social-Semantic Web 3.0) zu einem interaktiven sozio-technischen System. Die Grundlagen des Pragmatic Webs sind vielfältig und kommen dabei aus verschiedenen Gebieten, wie z. B. Language Action Perspective, Pragmatics, Semiotics, Agententheorien, etc. Überschneidungen gibt es auch mit dem neuen interdisziplinären Gebiet „Web Science“.

2.6 Zusammenfassung und Ausblick „Ubiquitous Pragmatic Web 4.0“

Die Entwicklung des Webs erfolgt in Stufen. Hervorgegangen aus der monolithischen Ära der Arbeitsplatzrechner war das erste Web 1.0 auf die Verknüpfung und Präsentation von Informationen für Menschen mit syntaktischen Formaten wie HTML angelegt. Mit dem Web 2.0 und dem zunehmenden Einzug von dynamischen Webanwendungen und -dienste wurde der Bedarf an maschinell bedeutsamer Semantik größer, was zum Semantic Web führte. Das Semantic Web versieht Webinformationen mit explizitem Wissen und einer (formalen) Semantik und verknüpft dieses semantische Wissen im Web, so dass es von Maschinen gefunden und für automatisierte Interpretationen genutzt werden kann. Speziellere Anwendungen semantischer Technologien führten in den letzten Jahren z. B. zum Corporate Semantic Web (Anwendung im Unternehmensumfeld [4]) und dem Social

Semantic Web (Anwendung in sozialen Interaktionen) oder auch semantischen Websdiensten (Semantic Web Services), sowie neuerdings der semantischen Realzeitdatenverarbeitung (wie dem Semantic Complex Event Processing (SCEP), d. h. der Erkennung komplexer Ereignisse mit Hilfe zusätzlicher semantischer Wissensbasen). Im Fokus des aktuellen Webs 3.0 steht dabei die Verbindung und Interaktion von Menschen, Diensten und Dingen mittels neuartiger Webschnittstellen und Werkzeuge, die semantisch unterstützt das Wissen des Webs nutzen können.

Stand 2014 ist das Web mitten in einem weiteren Paradigmenwechsel begriffen, von einem passiven World Wide Web, in dem Anbieter Web-Inhalte erzeugen und zum Abruf als HTML-Dokumente über das Internet anbieten, hin zu einem aktiven alles durchdringenden Pragmatic Web (**Ubiquitous Pragmatic Web 4.0** [5, 6]), welches intelligente Inhalte (**Corporate Smart Content** [11]) über pragmatische Internetdienste (z. B. „end-user programable intelligent agents“ [7, 8]) mit neuartigen Nutzerschnittstellen („smart devices“) und Dingen („smart things“) verbindet. Diese Entwicklung steht in engen Einklang mit der Vision des W3C für eine neuartige W3C „**Open Web Platform**“, welche sich anstelle des klassischen Browser-basierten Zugangs zum Web mit neuen Schnittstellen und deren Technologien (z. B. smart TV) beschäftigt. Abbildung 2.3 zeigt diese Entwicklung des Webs und den Paradigmenwechsel von einem passiven Web der Inhaltsproduzenten zu einem aktiven Web der Informationskonsumenten.

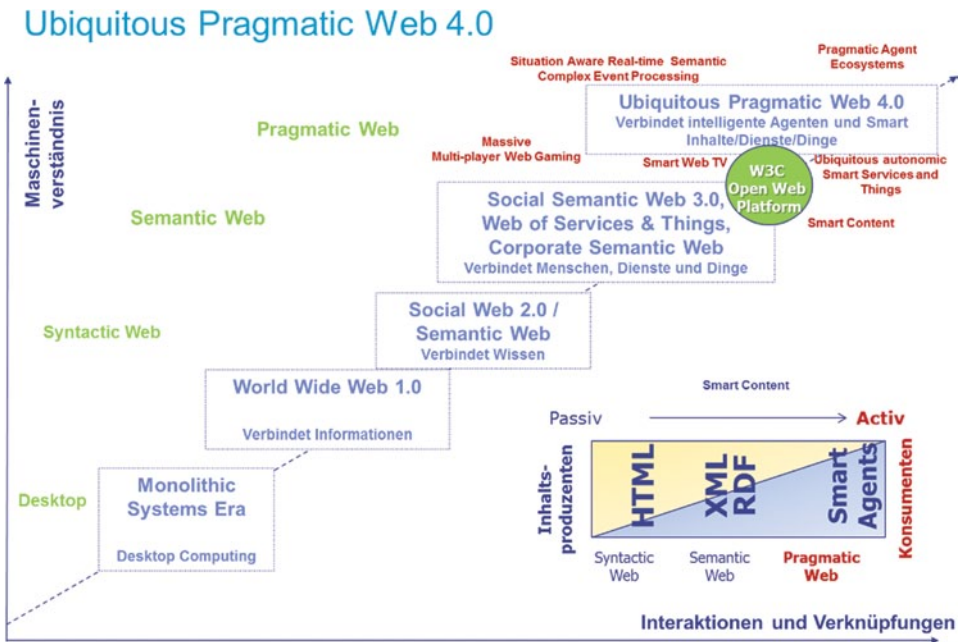


Abb. 2.3 Entwicklung zum allumfassenden Pragmatic Web 4.0. (Quelle: [6])

Von zentraler Bedeutung in diesem neuen interaktiven Web ist dabei der pragmatische Interaktionsrahmen, welcher das kontextuelle und modale Verständnis liefert oder auch die situativen Bedürfnisse und Verhaltensvorschriften vorgibt, wie z. B. Rechte und Pflichten. Erst dieser pragmatische Rahmen ermöglicht eine sinnvolle semantische Filterung und Transformation der Flut an oft multi-medialen Daten in relevante Informationen und Wissen, die in richtiger Form und zum richtigen Zeitpunkt für die Nutzer aufbereitet aktiv über das Internet übermittelt werden können. Beispiele kommen hier nicht nur aus der zunehmend interaktiv werdenden Web-Unterhaltungs- und Werbeindustrie (Stichwort z. B. Smart TV, Google Glass), sondern auch durch neuartige aktive Inhalte (**Smart Content**) und Dienste (Smart Services and Things), wie sie zum Beispiel im Bereich Online-Lehre und -Training in der Ausbildung oder auch im Bereich der aktiven Situationserkennung in einem Corporate Pragmatic Web nutzbar wären. Wie in Abb. 2.1 dargestellt, liegt dabei eine große Herausforderung an dieses „*Smarte Web*“ in der Erweiterung klassischer formaler Logiken, wie sie im Semantic Web genutzt werden, mit Logiken, die mit unscharfem Wissen umgehen und induktiv pragmatische „Querschlussfolgerungen“ erlernen können.

Literatur

1. Heese, Ralf, Gökhan Coskun, Markus Luczak-Rösch, Radoslaw Oldakowski, Adrian Paschke, Ralph Schäfermeier, und Olga Streibel. 2010. Corporate Semantic Web – Semantische Technologien in Unternehmen. *Datenbank-Spektrum* 10 (2): 73–79.
2. Blumenaur Andreas, Tassilo Pellegrini, und Adrian Paschke. 2010. Corporate Semantic Web – Der Einsatz von Semantic Webtechnologien im Unternehmen. HMD Praxis der Wirtschaftsinformatik 275:105–114. ISSN 1436–3011.
3. Paschke, Adrian, Gökhan Coskun, Ralf Heese, Markus Luczak-Rösch, Radoslaw Oldakowski, Ralph Schäfermeier, und Olga Streibel. 2010. *Corporate semantic web: Towards the deployment of semantic technologies in enterprises*, Hrsg. W. Du und F. Ensan, 105–131. New York: Springer.
4. Paschke, Adrian, et al. Corporate Semantic Web, InnoProfile project funded by the German Federal Ministry of Education and Research (BMBF) and the BMBF Innovation Initiative for the New German Länder – Entrepreneurial Regions. <http://www.corporate-semantic-web.de>. Zugegriffen: 15. Okt. 2014.
5. The Pragmatic Web Community. <http://www.pragmaticweb.info>. Zugegriffen: 15. Okt. 2014.
6. Weigand, Hans, und Adrian Paschke. 2012. The pragmatic web: Putting rules in context, 6th International Conference on Rules (RuleML 2012), Montpellier, France, August 27–31, 2012.
7. Paschke, Adrian, Harold Boley, Alexander Kozlenkov, und Benjamin Craig. 2007. Rule responder: RuleML-based agents for distributed collaboration on the pragmatic web, 2nd International Conference on the Pragmatic Web, Tilburg, Netherlands, 2007.
8. Paschke, Adrian, und Harold Boley. 2011. Rule responder: Rule-based agents for the semantic-pragmatic web, in Special Issue on Intelligent Distributed Computing in International Journal on Artificial Intelligence Tools (IJAIT), Vol. 20. Juni 2011.
9. O’Reilly, T. 2005. What is Web 2.0. Design patterns and business models for the next generation of software. <http://oreilly.com/web2/archive/what-is-web-20.html>. Zugegriffen: 13. Sept. 2014.

10. Breslin, John G., Alexandre Passant, und Stefan Decker. 2009. *The social semantic web*. 1 Aufl. New York: Springer.
11. Paschke, Adrian, et al. Corporate smart content. <http://www.corporate-smart-content.de>. Zugegriffen: 15. Okt. 2014.



Prof. Dr. rer. nat. Adrian Paschke ist seit 2008 Professor und Leiter der Arbeitsgruppe Corporate Semantic Web an der Freien Universität Berlin. Seine Forschung, u.a. im gleichnamigen BMBF Projekt „Corporate Semantic Web“, hat führend zur Entwicklung und Anwendung semantischer Technologien im Unternehmensumfeld beigetragen. In über 120 wissenschaftlichen Publikationen hat er maßgeblich die Entwicklung des Corporate Semantic Web und Pragmatic Web vorangetrieben, u.a. auch in der Standardisierung semantischer Technologien, wie im W3C Semantic Web Stack (RIF), RuleML (Reaction RuleML), OMG (API4KB), OASIS (LegalRuleML) sowie EPTS (EPTS-RA) und in Open Source Projekten wie DBPedia Deutsch und Prova. Daneben war er aktiv in der Semantic Web Gemeinde in einer Vielzahl an Konferenzorganisationen

(z. B. i-Semantics, ESWC, SWAT4LS, RuleML, ICPW, edBPM), Leitung technischer Standardisierungsgruppen, als Direktor von RuleML Inc. und Vizedirektor von STI Berlin, Gründer der Berlin Semantic Web Meetup Gruppe sowie in Forschungsausgründungen, wie der Klickfilm GmbH und Chariteam UG.



Dipl.-Inform. Ralph Schäfermeier ist Wissenschaftlicher Mitarbeiter an der Arbeitsgruppe Corporate Semantic Web der Freien Universität Berlin unter der Leitung von Prof. Dr. Adrian Paschke. Er studierte Informatik und Psychologie an der Technischen und der Freien Universität Berlin und schloss sein Studium an der Freien Universität ab. Er arbeitete als Softwareentwickler in zahlreichen Firmen und Projekten u. a. in den Bereichen Content Management, Credit Risk Management und Software for Life Science. Sein derzeitiger Forschungsschwerpunkt liegt im Bereich des Ontology Engineering, insbesondere in der aspekt-orientierten Ontologie-Entwicklung sowie generell in agilen Entwicklungsansätzen für Ontologien und Ontologie-Modularisierung. Frühere Arbeiten von Ralph

Schäfermeier beschäftigten sich mit dem Thema der automatisierten Wissensgewinnung durch die Beobachtung von Nutzern in ihrem kollaborativen Agitationskontext und im Bereich der kollaborativen Evolution von Wissen.