

XML, Standards und andere Aktivitäten zur Formierung des Semantic Web

Mario Jeckle
DaimlerChrysler Forschungszentrum Ulm
Wilhelm-Runge-Str. 11
89013 Ulm
mario.jeckle@daimlerchrysler.com
www.jeckle.de

Abstract: Die Vision eines *Semantischen Webs* als Grundlage einer nächsten Nutzungsgeneration des World Wide Webs nimmt derzeit breiten Raum in der fachlichen Diskussion ein [Be01a,Ge02,Ho02,Ma02].

Oftmals ist jedoch die Auffassung dessen, was diesen nächsten Entwicklungsschritt konstituiert, höchst disparat. Kennzeichnendes Merkmal aller Einlassungen zur Thematik scheint jedoch die Betonung der Bedeutung abgestimmter offener und daher breit unterstützter Standards, wie die des World Wide Web Konsortiums (W3C), zu sein.

Der vorliegende Betrag unternimmt den Versuch, die Grundidee des Semantischen Webs zu skizzieren und gleichzeitig einige der verfügbaren Basistechniken nebst ihrer präferierten Anwendungsgebiete einzuführen. Besonderes Augenmerk wird hierbei dem Technikgebiet der *Extensible Markup Language* [Br00] und der darauf aufbauenden Sprachstandards zuteil.

Den Abschluß bildet ein kurzer summarischer Überblick aktueller Aktivitäten und noch zu leistender Schritte zur Erreichung des durch die Vision formulierten Ziels.

1 Idee des Semantic Web

TIM BERNERS-LEE formuliert in [Be01a] die Zielsetzung des Semantischen Webs lakonisch als eine Erweiterung des gegenwärtigen Webs, in der Information über eine wohldefinierte Bedeutung verfügt, um eine bessere Zusammenarbeit zwischen Mensch und Computer zu ermöglichen. Unterstellt man der Formulierung *gegenwärtiges Web* die Bedeutung eines zusammenfassenden Begriffs als Gesamtheit der aktuell eingesetzten Techniken zum Auffinden, Darstellen, Übertragen und Präsentieren beliebiger Daten, so wird offenbar, daß der Autor – immerhin Direktor des World Wide Web Konsortiums und „Erfinder“ des größten Teils der gegenwärtig eingesetzten Techniken wie URL, HTTP und HTML –, diesen eine Verfehlung des Ziels einer zumindest gut zu nennenden Zusammenarbeit von Mensch und Maschine zuschreibt.

Im wesentlichen speist sich die formulierte Kritik aus dem bisher mit automatisierten technischen Mitteln nicht vollziehbaren Übergang der durch das weltweite Netz angebotenen

Daten zu Informationen im SHANNON'schen Sinne [Sh48]. Diese Interpretation bleibt für alle aus dem Web gewonnenen Daten dem kognitiven Prozeß des Empfängers überlassen.

Dasselbe gilt umgekehrt für die von einem menschlichen Nutzer zu Daten reduzierten Informationen, die an das Netz – etwa an Suchmaschinen – übermittelt werden. Das vorherrschende Codierungsformat für Daten beider Transferrichtungen ist hierbei nahezu ausschließlich in einer natürlichen Sprache formulierter Text.

In der Konsequenz dieses Interaktionsparadigmas ergeben sich Nutzungsmuster, die jedem WWW-Anwender vertraut sind:

- Datenproduktion¹: Zur Publikation beliebiger Inhalte im weltweiten Netz werden Textsprachen verwendet, deren überwiegendes Einsatzfeld die Festlegung graphischer Eigenschaften der anzuzeigenden Daten ist.

Um diese Daten für Dritte auffindbar abzulegen ist Kenntnis der Funktion sog. „Suchmaschinen“ notwendig.

- Datensuche: Recherchevorgänge im Web lassen sich in zwei Kategorien gliedern: Zunächst geführte Suchen durch manuell erzeugte und befüllte Verzeichnisse, die Netzseiten gemäß vorgegebener Kategorien rubrizieren,² zum anderen automatisiert mit aus den indizierten Seiten extrahierten textuellen Daten gefüllten Datenbanken.³

Während der erste Ansatz das Ergebnis eines manuell durchgeführten Interpretationsvorganges festschreibt läßt der zweite keinerlei Interaktion beim Aufbau des Datenbestandes zu.

Seitens des Anfragers offenbaren sich in der Praxis jedoch beide Ansätze zumeist als inadäquat. Unterdrückt der erste potentielle Treffer, wenn keine Einordnung in die angefragte Rubrik vorliegt, so liefern Umsetzungen des letzteren Ansatzes häufig eine Fülle von Ergebnisseiten, die zwar den gesuchten Begriff lexikalisch enthalten, hierbei jedoch den Auftretenskontext außer Acht lassen und die Ergebnismenge daher auch auf unpassende oder gar unerwünschte⁴ Inhalte unnötig aufblähen.

Um Suchmaschinen effizient zur Identifikation relevanter Ergebnisseiten heranziehen zu können ist die manuelle Auswertung der Suchergebnisse unabdingbar.

Zusätzlich offenbart sich die anwenderformulierte Anfrage der Suchmaschinen im zweiten Realisierungsmuster bzw. die Auswertung der natürlichsprachlich benannten Kategorien im ersten als inadäquat, da auch sie eines manuellen Nutzereingriffs bedarf. Dieser geschieht entweder durch direkte Formulierung der Anfrage als Zeichenkette oder durch Auswertung der angebotenen Kategorien, welche wieder durch verzeichnisdienstspezifische Benennungen identifiziert werden.

Um mit den aktuell vorherrschenden Mechanismen effiziente Suchen hinsichtlich des zu erbringenden Zeitaufwandes in Relation der als relevant erachteten Resultat-

¹Im Sinne einer Semantischen Abreicherung der Information zu Textdaten

²Dieser Gruppe sind Verzeichnisse wie „Yahoo“ und „dmoz“ zuzuordnen.

³Dieser Gruppe sind Verzeichnisse wie „Altavista“ und „Google“ zuzuordnen.

⁴Als ein besondere Form der Online-„Werbung“ ist seit einiger Zeit die gezielte Applizierung häufig gesuchter Begriffe zu beobachten, um möglichst viele fehlgeleitete Zugriffe auf eine Netzpräsenz zu erhalten.

seiten zu den tatsächlichen Suchergebnissen⁵ zu erzielen, ist daher die Anreicherung der durch Suchmaschinen verwalteten Basisinformation um zusätzliche deskriptive Anteile notwendig.

- **Datenkonsumption:** Die menschliche oder maschinelle Verarbeitung von Web-verfügbaren Inhalten kann ausschließlich auf den verfügbaren textuellen Inhalten erfolgen. Jegliche Bedeutung der empfangenen Daten muß durch eine manuelle Interpretation (re-)konstruiert werden, sofern dies überhaupt möglich ist.

Um aus dem Web bezogene Daten maschinell weiterverarbeiten zu können ist daher im Rahmen der Datenkonsumption die manuelle Anreicherung mit Kontext- und Anwendungssemantik notwendig.

Die skizzierten mehrfach notwendigen Interpretationsvorgänge kennzeichnen den aktuellen Umgang mit Web-verfügbaren Inhalten und motivieren gleichermaßen die Vision des Semantischen Webs.

Im Spiegel der diskutierten Nutzungsmuster ergibt sich die Zielsetzung der Semantic Web Initiative des World Wide Web Konsortium im Stile der durch BERNERS-LEE erhobenen Forderung: Techniken zu entwickeln und zu etablieren, welche die Interaktion zwischen menschlichem Nutzer und Computer vereinfachen.

2 Techniken zur Realisierung

Bereits die ersten publizierten Ideen zur technischen Umsetzung des Semantic Web (etwa in: [Be98]) offenbaren die Spannweite des Ansatzes. Sie reicht von der Basisvoraussetzung der maschinenverarbeitbaren Codifizierung aller Daten, über ihre Anreicherung mit der notwendigen Zusatzinformation in ebenfalls maschinenverarbeitbarer Darstellung bis hin zur Integration von Ontologien und Sicherheitsaspekten wie digitaler Signatur und Verschlüsselung.

Durch die umfassend erhobene Forderung nach Maschinenverarbeitbarkeit läßt sich aus der Grundidee zunächst die Forderung nach der Schaffung eines Formats ableiten, welches sowohl mächtig genug ist, die geforderte Datendarstellung zu leisten, als auch gleichzeitig hinreichend formalisiert ist, um durch Computer prozessiert zu werden. Überdies läßt sich BERNERS-LEES Vision in diesem Kontext dahingehend interpretieren, daß die zur Maschinenverarbeitung erzeugten Daten auch für Menschen lesbar sein sollen.

Um diesem Grundforderungskatalog zu genügen wurde die Sprache *Extensible Markup Language* (XML) [Br00], als Untermenge der Sprache *Standardized Generalized Markup Language* (SGML) [In86], definiert. XML realisiert eine Unicode [Un00] basierte Metasprache zur Definition beliebiger Vokabulare⁶. Hierdurch werden zwei wesentliche,

⁵Als solche seien in diesem Kontext diejenigen Ergebnisse einer Suche bezeichnet, die zu Seiten führen die einen Informationswert, gemessen an Neuheit und Verwertbarkeit für den Anfragenden, darstellen.

⁶Als Differenzierung der *Sprache* vom reinen *Vokabular* sei im vorliegenden Papier die Existenz einer Semantikbeschreibung zusätzlich zur reinen Syntax maßgeblich.

für den Anwendungskontext des Semantischen Webs erfolgskritische Freiheitsgrade erreicht: Zum einen gestattet die Anlage der XML als generische Metasprache die freie Erzeugung von Inhalten, deren syntaktische Struktur ein einheitliches – und somit maschinenverarbeitbares – Metamodell aufweist und gleichzeitig die anwendungsspezifische Modellbildung auf der Problemebene zuläßt. Zum anderen resultiert aus der Heranziehung des Unicode Standards, der als Zeichensatz die meisten der gebräuchlichen Alphabete und Syboldarstellungen⁷ enthält, die Möglichkeit die abzulegenden Daten in ihrer Originalsprache unübersetzt und -transkribiert abzulegen.

Aufgrund dieser Flexibilität findet XML als Ausgangspunkt weiterer Standards, welche die konstituierenden Architekturelemente des Semantic Web bilden, breiten Einsatz.

2.1 W3Cs XML-Schema und XML-Standards der zweiten Generation

Die vergleichsweise einfache Erzeugung neuer Vokabulare ausgehend vom Urstandard XML führt in der kurzen Vergangenheit seit seiner Veröffentlichung zu einer Flut neuer Formate, die sich dieser Metasprache bedienen. So wurden neben der Reformulierung zuvor — aus XML-Sicht — proprietär gestalteter Syntaxen auch zunehmend neu entwickelte Serialisierungs- und Austauschformate als XML-Sprachen realisiert.

Ein wesentliches Charakteristikum bei diesen XML-basierten Formaten bildet ihr Einsatz in datenorientierten Umgebungen. Dieses Anwendungsfeld bildet den Motivationsstrang, der als ursächlich für die Entwicklung der bis dato wirkungsmächtigsten XML-Sprache anzusehen ist — W3Cs XML-Schema.

War ursprünglich XML, ebenso wie der Vorgängerstandard SGML, auf die präsentationsorientierten Anwendungsfelder des Verlagswesens und der technischen Dokumentation ausgelegt, so wurde durch die XML ein Anwendungsfeld mit gänzlich verändertem Anforderungsprofil erschlossen, in dem die Nachteile des „SGML-Erbes“ deutlich zu Tage treten:

- Unzureichende Datentypunterstützung.
Für die Datencodierung stehen ausschließlich auf einfache uninterpretierte Zeichenketten zurückführbare Datentypen zur Verfügung.
- Unzureichende Strukturierungsunterstützung.
Die Inhalte eines XML-Datenstroms sind strikt hierarchisch organisiert und lassen weder Querbezüge noch aussagekräftige konsistenzgarantierende Zwangsbedingungen zu.
- Keine Wiederverwendungsunterstützung.
Die Nutzung bereits erzeugter Strukturen in bestehenden XML-Grammatiken ähnlicher Einsatzkontexte ist praktisch kaum möglich.

⁷In der aktuellen Fassung sind dies weit über 2 Milliarden Zeichen

- Starres Typsystem.

Die Erweiterung des vorgegebenen Typsystems durch den Anwender ist nicht vorgesehen und findet auch durch existierende Sprachmechanismen keine adäquate Unterstützung.

- Keine Unterstützung von Namensräumen.

Namensräume [Br99] für die Familie der durch die Vokabulardefinition festgelegten Dokumente können nicht formuliert werden.⁸

- Grammatiksyntax ist keine XML-Sprache.

Die von SGML übernommene Grammatiksyntax basiert auf einer separaten Metasyntax, die durch Entwickler gesondert zu erlernen und durch Werkzeugen gesondert zu implementieren ist.

Angewendet auf die Vision des Semantic Web erweist sich XML im Lichte der dargestellten Unzulänglichkeiten in seiner Grundform als inadäquater Lösungsbaustein. Unterstützt sie doch nur unzureichend die Anforderungen an die maschinelle Verarbeitung der Inhalte und erfordert aufwendige Interpretationsschritte zur (Wieder-)Gewinnung der applikatonsseitig vor dem Encodierungsvorgang verfügbaren Daten.

Um mit den diskutierten Einschränkungen auszuräumen, und in diesem Zuge die Nutzung XML-basierter Vokabulare im datenorientierten Anwendungsumfeld zu befördern, wurde durch eine Arbeitsgruppe des World Wide Web Konsortiums die Sprache *XML-Schema* [Be01,Bi01] erarbeitet.

Sie formuliert ein XML-Vokabular gemeinsam mit einer englischen Semantikbeschreibung⁹ zur Erstellung beliebiger XML-Vokabulare. Gleichzeitig führt [Bi01] ein erweitertes und durch den Anwender erweiterbares Typsystem ein.

W3Cs XML-Schema findet heute breite Anwendung zur Definition neuer XML-Vokabulare. Zusätzlich markiert dieser Standard den Übergang zur zweiten Generation der XML-Standards. Im Gefolge entstehen durch das W3C standardisierte Sprachen, welche die erweiterten Strukturierungsmöglichkeiten nutzen, um XML-codierte Inhalte datenorientiert verarbeiten zu können. Zusätzlich wird durch die geschaffene Selbstbeschreibung von XML die Möglichkeit der Verarbeitung von XML-Vokabularen, sowohl durch XML-Sprachen wie XPath [C199] oder XSLT [C199a] als auch durch generische XML-Programmierschnittstellen, geschaffen.

2.2 Metadaten

Metadaten — im einfachsten Wortsinne als *Daten über Daten* — zu XML codierten Inhalten lassen sich prinzipiell in frei wählbaren Formaten verfassen und ablegen. Vielmehr

⁸Die hierfür oft angeführten Hilfskonstrukte spiegeln weder die volle Mächtigkeit, noch die korrekte Semantik der Namensräume wieder.

⁹Eine kommentierte deutsche Übersetzung ist mit [Mi02] und unter <http://www.jeckle.de/xsd-de/> verfügbar.

noch, im weitesten Sinne erfüllt jegliche Art eines deskriptiven Datums zu einem gegebenen Datum die eingeführte Sinngebung.

Wesentlich für die Vision des Semantic Web ist jedoch nicht die Metadatenexistenz an sich, sondern der Wunsch nach Maschinenverarbeitbarkeit zur Unterstützung der Benutzerinteraktion mit den durch die Metadaten beschriebenen Daten.

Aus diesem Wunsche heraus wird die enge Verbindung des Technikgebietes XML mit der Idee des Semantic Web offenbar. Eignet sich doch die selbst als Metasprache angelegte XML in besonderer Weise zur Formalisierung der für das Semantic Web existenznotwendigen Metainformation; wenngleich diese Formalisierung selbst nur auf der deskriptiven Metaebene der syntaktischen Beschreibung anzusiedeln ist, wie einschränkend angemerkt sei. Vielmehr noch, verschiebt dieser — an sich naheliegende und valide — Formalisierungsansatz die notwendige Semantische Vereinheitlichung zu Gunsten der durch die XML erfolgten Syntaktischen auf die Meta-Metaebene. Durch die Selbstbeschreibungsfähigkeit der XML kollidiert jedoch auf dieser Ebene der Semantikbegriff XML-Schemas mit der dort zu etablierenden Semantikbeschreibung der durch das Schema definierten Sprache. Gänzlich unlösbar wird das Ansiedelungsproblem durch die Organisation des XML-Schemastandards als wiederum selbstbeschreibende Sprache, die den XML-Schemastandard selbst zu seiner syntaktischen und Semantischen Beschreibung heranzieht ...

Zu dieser syntaktisch induzierten Semantikansiedelungsproblematik tritt eine Semantisch induzierte Syntaxansiedelungsproblematik bei der strukturellen Definition der deskriptiven Metainformation selbst. Konkret stellt sich die Frage der zu wählenden Abstraktionsebene bei der Sprachformulierung einer allgemein einsetzbaren Metasprache zur Codierung beschreibender Daten zu XML codierten Daten.

Wird der Abstraktionsgrad zu hoch gewählt, so geht dies zwar mit einer prinzipiellen Verbreiterung des Einsatzgebietes, d.h. einer größeren Menge beschreibbarer Entitäten¹⁰ einher, schränkt jedoch in gleicher Weise die Spezifität der so formulierbaren Aussagen ein.

Wird der Abstraktionsgrad hingegen möglichst niedrig gewählt, so gewinnt man zwar einerseits ausdetaillierte Entitätsbeschreibungen, diese jedoch nur für sehr wenige Entitäten.

Vor dem Hintergrund dieser Problematik finden im Kontext des Semantic Web zwei Ansätze Verwendung, die in ihrer Kombination den Zielsetzungen beider konkurrierender Ansprüche zu genügen suchen.

So definiert das *Resource Description Framework* (RDF) [La99] ein Vokabular zur Darstellung beliebiger Aussage über *Ressourcen*. Zur Beschreibung von Syntax kann für RDF XML-Schema Anwendung finden.

Zur Semantikdefinition greift RDF jedoch nicht auf XML-Schema zurück, sondern die eigens zum Zwecke der RDF-Beschreibung entwickelte Metasprache *RDF-Schema* [Br02], welche wiederum selbstbeschreibend organisiert ist. Dieser Schritt fundiert die RDF zugeordnete Rolle eines formalisierten Metadatendefinitions vokabulars gegenüber XML-Schema, dessen Einsatzzweck in der syntaktischen Beschreibung beliebiger Vokabulare und der Semantikdefinition der XML-Schemasprache erschöpft ist.

¹⁰Der Begriff sei hier im Sinne eines *realen oder gedachten Gegenstandes der Anschauung* gebraucht.

RDF selbst bietet eine einfache XML-Syntax zur Formulierung von Aussagen über Ressourcen an. Der Begriff der *Ressource* ist dabei eng an den der eindeutigen Identifizierbarkeit gekoppelt. Diese wird hierzu mit der Technik der *Universal Resource Identification* (URI) [Be98] synonymisiert. Inhaltlich orientieren sich die mittels RDF formulierbaren Aussagen an der Satzstruktur der englischen Sprache und gestatten Zuordnungen von prädikativ angebotenen Werten (Objekt) zu Ressourcen (Subjekt).

Auf der Basis des RDF-Standards ist es somit möglich, in ihrem Detaillierungsgrad frei variierende deskriptive Daten zu einer beliebigen Web Ressource so abzulegen, daß sie auf der Basis von XML maschinell weiterverarbeitet werden können. Syntax und Struktursemantik werden hierbei durch die W3C-Empfehlungsdokumente spezifiziert. Ungeklärt ist bisher lediglich noch die Semantische Charakterisierung der deskriptiven Dateninhalte selbst. Durch die Etablierung des Resource Description Frameworks wird zwar ein wesentlicher Schritt hin zu einer semantikbasierten Interoperabilität unternommen, jedoch kann wirkliche Semantik auch durch RDF nicht ausgedrückt werden. Im Grunde handelt es sich bei RDF um einen sehr flexibel einsetzbaren und leistungsfähigen Mechanismus zur freien Ressourcennotation, welcher letztlich die deskriptiven Daten auf eine textuelle Repräsentation in einer natürlichen Sprache zurückführt.

Aus dieser Mimik ergeben sich die nachgerade klassischen Synonym-Homonym-Probleme hinsichtlich der begrifflichen Eindeutigkeit von Elementen einer natürlichen Sprache bei ihrer maschinellen Verarbeitung. Dieser Herausforderung widmet sich die W3C-Arbeitsgruppe *Web Ontologies* [We01], welche jedoch bisher noch keine Resultate veröffentlichte.

2.3 Web-Dienste

Waren die bisher umrissenen Techniken und Standards primär zur Realisierung der Vision des Semantic Web im Kontext des aktuell vorherrschenden Interaktionsparadigmas zwischen Mensch und Computer angesiedelt, so sucht die Idee der Web-Dienste (*Web Services*) die Tür in eine neue Generation der Internetnutzung im allgemeinen aufzustoßen.

Web-Dienste sollen aufbauend auf den Techniken des sichtbaren, leicht zu bedienenden und grundsätzlich auf menschliche Interaktion zentrierten Internets der zweiten Generation einen evolutionären Schritt zu einer dritten Generation vollziehen, in welcher sich die Netzverfügbarkeit zur Allgegenwart auswächst, wobei gleichzeitig die Sichtbarkeit — und damit die bewußte Wahrnehmung der Interaktion — immer mehr in den Hintergrund treten und letztendlich vollkommen verschwinden soll.

Technisch gesehen handelt es sich bei Web-Diensten lediglich um die Möglichkeit, entfernte Funktionalität, unbeachtlich ihrer internen Realisierung als einzelne Funktionen, Methoden oder komplexe Dienstleistungen, zu nutzen. Hierbei kommen plattform- und programmiersprachliche Ansätze, wie XML in Kombination mit Elementen des Internetprotokollstapels zum Einsatz, welche sich durch eine größtmögliche Generizität bei gleichzeitig hohem Interoperabilitätspotential, begründet durch die beiden Erklärungskomponenten: Verbreitung und Einfachheit, auszeichnen.

Im Sinne des Semantischen Webs versprechen Web-Dienste den Brückenschlag hin zur automatisierten Verarbeitung der Metadaten, wobei sich diese Verarbeitung selbst wieder auf Vokabulare und Sprachen des Semantic Web stützt bzw. sich dieser bedient.

Wesentlicher Konstituent der im Entstehen begriffenen Standards aus diesem Umfeld ist das SOAP-Protokoll, welches derzeit durch eine Arbeitsgruppe des World Wide Web Konsortiums erarbeitet wird. Es gestattet den Aufruf entfernter Methoden (*Remote Procedure Call*) sowie den Austausch dokumentartiger Daten ausgehend von einer durchgängigen XML-Codierung der übertragenen Inhalte. Neben der daher naheliegenden Übertragung maschinenlesbarer Metadaten getrennt von der präsentationsorientierten Aufbereitung für den Menschen, die üblicherweise durch einen Internet Browser verwirklicht wird, fügen sich Web-Dienste zusätzlich auf eine zweite Weise in die Vision des Semantic Web ein.

Jeder Web-Dienst selbst repräsentiert eine identifizierbare Netzressource im Sinne des Resource Description Framework. In der Tat stellt die Beschreibung und Verfügbarkeit von Metadaten zu angebotenen Web-Diensten den zentralen kritischen Aspekt des Web Service Gedankens dar. Aus diesem Grunde befindet sich derzeit ein mit RDF abgestimmter Sprachstandard zur funktionalen Beschreibung dieser Dienste in der Gremiendiskussion. Fernziel dieser Entwicklung ist die Verwaltung aller beschreibenden Daten in einem verteilt organisierten Metadatenverzeichnis, dessen Dienste selbst durch Metadaten im verteilten Format beschrieben sind und in der Konsequenz das vollautomatisierte Auffinden von Diensten anhand einer Metadaten-basierten Anfrage gestatten.

3 Aktueller Status und Ausblick

War das Semantische Web zum Zeitpunkt seiner ersten Erwähnung noch eine reine Vision, ohne jegliche technische Fundierung, so hat die Idee in den vergangenen Jahren stetig durch neue Standards und ihre Umsetzung an Form gewonnen.

Den wesentlichen Beitrag leistet die generische Metasprache XML, welche ursprünglich lediglich als HTML-Erweiterung konzipiert wurde. Dennoch verhalf ihr ihre Generalität zur Erschließung immer weiterer Einsatzfelder. Ein zentrales Element hierbei bildet die Möglichkeit, XML-basierte Sprachen frei zu kombinieren und das so entstandene Technikgebiet sukzessive um neue Bausteine anzureichern. Der wirkungsmächtigste dieser XML-Sekundärstandards bildet unumstritten W3Cs XML-Schemasprache, welche durch ihre erweiterten Strukturmechanismen gemeinsam mit der Schaffung eines leistungsfähigen Typsystems für XML den Übergang zur datenorientierten Verarbeitung von Webinhalten ermöglichte.

Auf dieser Basis entsteht derzeit die technische Realisierung der Basisinfrastruktur des Semantic Web in Gestalt des Resource Description Frameworks. Gegenwärtig entwerfen verschiedene Arbeitsgruppen des World Wide Web Konsortium weitere Standards, die auf diesem Ansatz aufsetzen. Hierzu zählen einerseits die Sprachen zur formalen Darstellung deskriptiver Metaaussagen, wie sie durch RDF-Schema und die Web-Ontologie-Initiative vorangetrieben werden.

Andererseits ist bereits durch Aktivitäten, die auf Prüfung und Gewährleistung von Verläss-

lich- und Vertraulichkeit der übermittelten (Meta-)Daten zielen, bereits der nächste evolutionäre Schritt des Semantischen Webs abzusehen, der sich in einem globalen Netz automatisiert ermittelbarer rechtlich bindender Partnerschaften manifestieren soll.

Literatur

- [Be01] Beech, D.; Lawrence, S.; Maloney, M.; Mendelsohn, N.; Thompson, H. S. (eds.): XML Schema Part 1: Structures. World Wide Web Konsortium, Boston, USA, 2001. W3C Recommendation. URL: <http://www.w3.org/TR/2001/REC-xmlschema-1-20010502/>.
- [Be01a] Berners-Lee, T.; Hendler, J.; Lassila, O.: The Semantic Web. In: Scientific American, May, 2001, URL: <http://www.scientificamerican.com/2001/0501issue/0501berners-lee.html>.
- [Be98] Berners-Lee, T.: Semantic Web Road Map. World Wide Web Konsortium, Boston, USA, 1998. URL: <http://www.w3.org/DesignIssues/Semantic.html>.
- [Be98] Berners-Lee, T.; Fielding, R.; Masinter, L.: Uniform Resource Identifiers (URI): Generic Syntax. Internet Engineering Task Force, Reston, 1998.
- [Bi01] Biron, P. V.; Malhotra, A. (eds.): XML Schema Part 2: Datatypes. World Wide Web Konsortium, Boston, USA, 2001. W3C Recommendation. URL: <http://www.w3.org/TR/2001/REC-xmlschema-2-20010502/>.
- [Br99] Bray, T.; Hollander, D.; Layman, A. (editors): Namespaces in XML. World Wide Web Konsortium, Boston, USA, January 1999. W3C Recommendation. URL: www.w3.org/TR/REC-xml-names.
- [Br00] Bray, T.; Paoli, J.; Sperberg-McQueen, C. M.; Maler, E. (eds.): Extensible Markup Language (XML) 1.0 (Second Edition). World Wide Web Konsortium, Boston, USA, October 2000. W3C Recommendation. URL: <http://www.w3.org/TR/2000/REC-xml-20001006>.
- [Br02] Brickley, D.; Guha, R. V. (eds.): RDF Vocabulary Description Language 1.0: RDF Schema. World Wide Web Konsortium, Boston, USA, April, 2002. W3C Working Draft. URL: <http://www.w3.org/TR/rdf-schema>
- [Cl99] Clark, J.; DeRose, S. (eds.): XML Path Language (XPath). World Wide Web Konsortium, Boston, USA, November 1999. W3C Recommendation. URL: <http://www.w3.org/TR/1999/REC-xpath-19991116>.
- [Cl99a] Clark, J. (editor): XSL Transformations (XSLT) Version 1.0. World Wide Web Konsortium, 1999. W3C Recommendation. URL: <http://www.w3c.org/TR/xslt>.
- [Ge02] Geroimenko, V.; Chen, C. (eds.): Visualizing the Semantic Web. Springer, London, Berlin, Heidelberg, et al., 2002.
- [Ho02] Horrocks, I.; Hendler, J. (eds.): The Semantic Web. Proceedings of the First International Semantic Web Conference, Sardinia, Italy, June 9–12, Lecture Notes in Computer Science Vol. 2342, 2002.
- [In86] International Organization for Standardization (editor): Information processing – Text and Online Systems – Standard Generalized Markup Language (SGML). International Organization for Standardization, 1986. ISO 8879:1986, Geneva, CH.

- [La99] Lassila, O.; Swick, R. R. (eds.): Resource Description Framework (RDF) Model and Syntax Specification. World Wide Web Consortium, Boston, USA, February, 1999. W3C Recommendation. URL: <http://www.w3.org/TR/REC-rdf-syntax>.
- [Ma02] Maedche, A.: Ontology Learning for the Semantic Web. Kluwer Academic, Boston, London, 2002.
- [Mi02] Mintert, S. (ed.): XML & Co. Addison-Wesley, München, 2002.
- [Sh48] Shannon, C. E.: A Mathematical Theory of Communication. Bell Systems Technical Journal, Vol. 27, Jul/Oct, 1948, pp. 379–423, 623–656.
- [Un00] Aliprand, J. (ed.): The Unicode Standard 3.0. Addison-Wesley, Bonn, Reading, 2000.
- [W301] World Wide Web Consortium (ed.): Semantic Web. Activity Overview, World Wide Web Consortium, Boston, USA, 2001. URL: <http://www.w3.org/2001/sw>.
- [We01] World Wide Web Consortium (ed.): Web-Ontology (WebOnt) Working Group. Working Group Homepage, World Wide Web Consortium, Boston, USA, 2001. URL: <http://www.w3.org/2001/sw/WebOnt/>.