

Metadaten als Bausteine des Semantic Web

The »semantic Web« represents one of the most important current trends in the continuous development of the World Wide Web (WWW). It is the ambitious vision of this next generation of the WWW to achieve a new quality in content and service provision through semantic enrichment of information, thereby opening up completely new opportunities for employing the Web. Important goals in the development of the Semantic Web include improved support for the cooperation between humans and computers and intelligent assistance for the execution of tasks within shared distributed information environments.

The clue to achieving these goals is the enrichment of data in the Web with metadata, which embeds these data in a semantic context. Such contextual information is interpreted by software applications and used for information filtering, refining queries and providing intelligent assistance. The design of a suitable format for the modelling and description of the context imposes a considerable challenge, since it is necessary to facilitate an automatic, global interpretation without recourse to a universal semantic schema. Establishing a consensus on such a universally valid schema is not possible within the framework of an extensive, heterogeneous and self-governing context as the World Wide Web.

Das »Semantic Web« bildet einen der wichtigsten, aktuellen Trends in der Weiterentwicklung des World Wide Web. Ehrgeizige Vision dieser nächsten Generation des WWW ist es, durch semantische Anreicherung von Information eine neue Qualität in der Bereitstellung von Inhalten und Diensten zu erreichen und vollständig neue Anwendungsmöglichkeiten für das Web zu eröffnen. Wichtige Ziele der Entwicklung des Semantic Web sind dabei die verbesserte Unterstützung von Kooperation zwischen Menschen und Computern und die intelligente Assistenz bei der Durchführung von Aufgaben in kooperativen verteilten Informationsumgebungen.

Schlüssel zur Erreichung dieser Ziele sind die Anreicherung von Daten im Web mit Metadaten, welche diese Daten in einen semantischen Kontext einbetten. Diese Kontextinformation wird durch Software-Anwendungen interpretiert und zur Informationsfilterung, Verfeinerung von Anfragen und zur Bereitstellung intelligenter Assistenten verwendet. Eine große Herausforderung stellt dabei die geeignete Modellierung und Beschreibung des Kontexts dar. Diese muss eine automatische, globale Interpretation ermöglichen, ohne dass auf ein allgemeingültiges semantisches Beschreibungsschema zurückgegriffen werden kann. Die Vereinbarung eines solchen allgemeingültigen Schemas ist in einem derart umfangreichen, heterogenen und autonomen Rahmen, wie ihn das WWW darstellt, nicht möglich.

DIE NÄCHSTE GENERATION DES WORLD WIDE WEB

Das World Wide Web und die übrigen Dienste im Internet, wie z.B. E-Mail, haben seit den Anfängen des Internets einen erstaunlichen Erfolg zu verzeichnen. Für viele Menschen ist die Nutzung des Internets aus ihrem täglichen Leben kaum mehr wegzudenken. Es ist fester Bestandteil des beruflichen und privaten Alltags geworden. Wir verwenden das Web, um Fahrpläne nachzuschlagen, Bankgeschäfte zu erledigen, Reisen zu buchen, Nachrichten zu lesen, Information zu recherchieren, Nachrichten auszutauschen, um nur einige Beispiele zu nennen.

Das World Wide Web (WWW) und die zugrunde liegende Internet-Technologie sind nicht statisch, sondern unterliegen der ständigen Weiterentwicklung in Bezug auf die verwendeten Technologien, Formate und die eingesetzte Software. Aktuelle Beispiele sind die Einführung von XML und die stetige Verbesserung der verwendeten Suchmaschinen. Diese Entwicklungen werden zum Teil durch das Aufkommen neuer Ansätze, Dienste und Produkte in der IT-Forschung, Entwicklung und Industrie getrieben, teilweise aber auch durch das immense Wachstum der im WWW verfügbaren Information und der ständig steigenden Benutzerzahlen notwendig gemacht.

Neben dieser kontinuierlichen Entwicklung gibt es auch weiter gehende Aktivitäten, deren Ziel es ist, eine neue Qualität im Leistungsangebot des World Wide Web zu erreichen. Aktuell gibt es in diesem Bereich zwei zentrale Trends:

► Das Paradigma der **Web Services** zielt auf eine Verwendung des »Web« als Programmierschnittstelle. Idee ist es dabei, Dienste im Internet verfügbar zu machen, die dynamisch als Komponenten in Softwaresysteme eingebunden werden können. Dabei werden die Kommunikationsprotokolle des Internets und eine Reihe XML-basierter Standards zur Dienstbeschreibung und zum Informationsaustausch verwendet (SOAP, WSDL, etc.). Die Entwicklung der Web Service Technologie und der benötigten Standards sowie der Aufbau der Infrastruktur für die Dienstvermittlung wird in starkem Maße von der IT-Industrie vorangetrieben.

► Dem **Semantic Web** liegt eine Anreicherung der im WWW angebotenen Information mit semantischer Information (Metadaten) zugrunde, welche die (teilweise) Interpretation der Information durch Software erlaubt und damit zahlreiche neuartige Dienste für den Nutzer des Semantic Web möglich macht.

Die Vision des Semantic Web muss als sehr ambitioniert bezeichnet werden. Sowohl die Inhaltsanbieter als auch die Technologieentwickler stehen in diesem Zusammenhang noch vor zahlreichen großen Herausforderungen.

Vision des Semantic Web

»The *Semantic Web* is an extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation.« (Berners-Lee et al. 2001)
Diese informelle Definition geben Tim Berners-Lee,



Claudia Niederée

dynamische Einbindung von Webdiensten in Softwaresysteme

Anreicherung von Information im Web mit Metadaten

**gezielte Informationssuche
im Semantic Web**

James Hendler und Ora Lassila in ihrem Aufsatz im *Scientific American*, der die Vision des Semantic Web illustriert. (»Das Semantic Web ist eine Erweiterung des aktuellen Web, welche die Kooperation zwischen Menschen und Computern erleichtern soll, indem Information eine klar definierte Bedeutung zugeordnet wird.«)

Die Vision des Semantic Web lässt sich am besten verstehen, wenn man von der typischen Situation bei der Suche nach Information im heutigen Web ausgeht [siehe Abbildung 1]: Ein Informationskonsument benötigt für eine bestimmte Aufgabe eine Information (oder einen Dienstanbieter) und begibt sich auf der Suche nach dieser Information in das WWW. Dem individuellen Informationsbedarf steht eine riesige Menge an Informationsressourcen (das WWW wird aktuell auf mehr als 3 Milliarden Dokumente ge-

schätzt) gegenüber, die von einer breiten Palette von Informations- und Dienst Anbietern zur Verfügung gestellt werden.

Typischer Einstiegspunkt in die Informationssuche im WWW sind neben speziellen Informationsportalen und persönlichen Lesezeichen (*Bookmarks*) die Internet-Suchmaschinen, die auf der Grundlage von vorausberechneten Indexstrukturen eine effiziente Suche im WWW ermöglichen. Obwohl Internet-Suchmaschinen wie z. B. Google (www.google.de) inzwischen eine hohe Qualität in Bezug auf ihr Suchverhalten besitzen, erhält der Benutzer immer noch eine große Menge an Suchergebnissen. Diese enthält auch viele Ressourcen, die keinerlei oder wenig Relevanz in Bezug auf die gesuchte Information besitzen. Das Abarbeiten eines solchen Suchergebnisses ist zeitaufwendig und bedarf gegebenenfalls auch der mehrfachen Umformulierung und Verfeinerung der Anfrage. Die Reduktion dieses Aufwands durch eine semantische Erfassung sowohl der Anfrage als auch der angebotenen Informationsressourcen ist eines der zentralen Ziele des Semantic Web.

Im Semantic Web gibt es zwischen dem Informationskonsumenten und dem Informationsanbieter eine erweiterte Dienst- und Informationsinfrastruktur [siehe Abbildung 2]. Dabei handelt es sich zum einen um Software-Programme, die so genannten Agenten, die vom Benutzer Aufträge entgegennehmen und diese autonom für ihn ausführen (Hendler 2001). Um ihre Aufgaben durchzuführen, können Agenten unterschiedliche Dienstleister, andere Agenten und auch andere Benutzer kontaktieren. Sie können verhandeln, planen und bis zu einem gewissen Grad auch Entscheidungen treffen. Man spricht deshalb auch von »intelligenten« Agenten, wobei zu beachten ist, dass Software im engeren Sinne nicht intelligent ist. Die Funktionsweise der Agenten besteht im Wesentlichen im Ausführen von definierten oder erlernten Regeln.

Um ihre Aufgaben erfüllen zu können, müssen die Agenten die Information nicht nur lesen, sondern auch interpretieren können. Zu diesem Zweck werden Informationsressourcen annotiert und zwar mit Information, die ihnen Semantik zuordnet. Neben den im Web angebotenen Informationsressourcen müssen auch die Informationen über den Benutzer, wie z. B. seine Präferenzen und Interessen, in semantisch annotierter Form vorliegen. Daraus ergibt sich der zweite Baustein des Semantic Web: die semantische Annotation von Informationsressourcen.

Der dritte Baustein des Semantic Web hat viel mit der Akzeptanz des Ansatzes zu tun. Wenn Agenten im Web im Namen des Benutzers Entscheidungen tref-

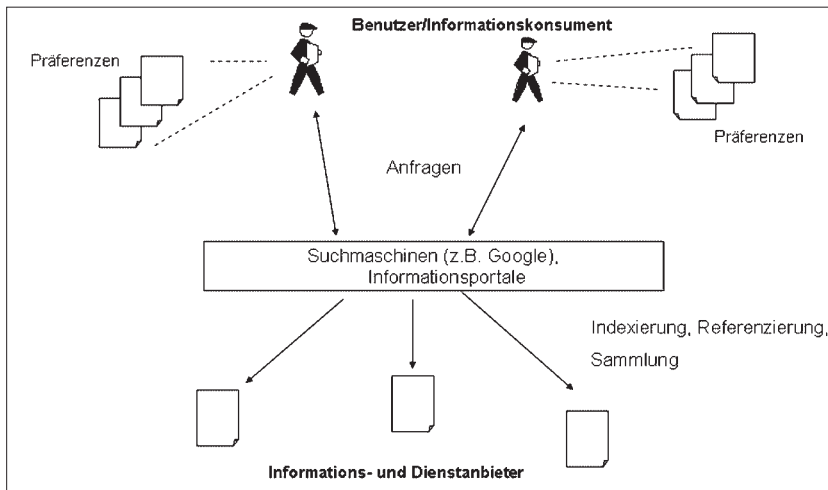


Abb. 1: Das World Wide Web – heute

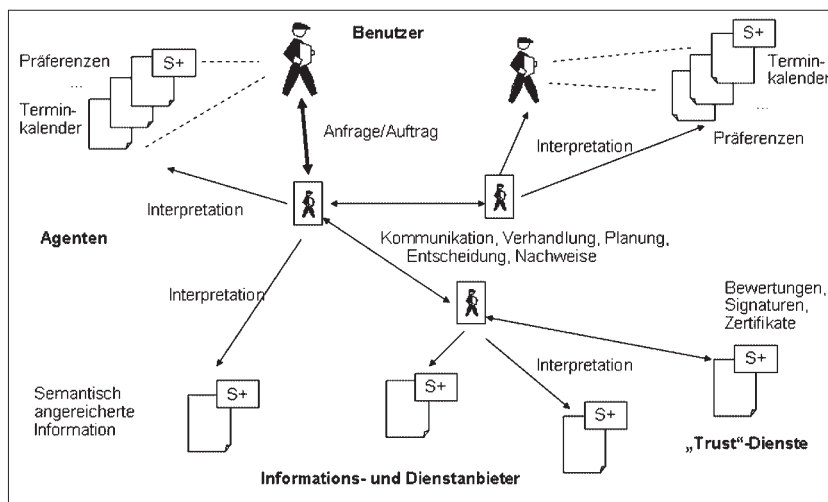


Abb. 2: Vision des Semantic Web

fen, stellt sich die Frage nach Verantwortlichkeiten und Vertrauen. So genannte »Trust«-Dienste bieten eine breite Palette von Maßnahmen an, die das Vertrauen des Benutzers in das Dienstangebot des Semantic Web erhöhen sollen. Dazu zählen digitale Unterschriften, Wasserzeichen, Nachweise für Entscheidungen, unterschiedliche Arten von Zertifikaten, etc.

Ein Agent im Semantic Web kann z. B. unter Berücksichtigung der persönlichen Präferenzen, Eigenschaften und Interessen des Benutzers einen Konferenzbesuch organisieren. Dabei kann er sowohl das Hotel buchen und bei der Auswahl eines Menüs die Präferenzen des Konferenzbesuchers berücksichtigen als auch relevante Konferenzpräsentationen und Workshops vorschlagen und Treffen mit anderen Wissenschaftlern gemäß der Forschungsinteressen des Konferenzbesuchers organisieren.

Beiträge zur Realisierung des Semantic Web

Die Umsetzung der ambitionierten Vision des Semantic Web erfordert Beiträge und Aktivitäten in unterschiedlichen Bereichen. Zur Anreicherung der Informationsressourcen mit semantischen Metadaten ergeben sich die folgenden Anforderungen:

- Es müssen **Modelle** und **Formate** für Metadaten zur semantischen Annotation entwickelt werden. Um einen weltweiten Austausch solcher Metadaten effizient zu ermöglichen, ist außerdem eine Standardisierung der Formate notwendig.
- Es werden **Methoden** und **Software-Werkzeuge** zur tatsächlichen Durchführung der semantischen Annotation benötigt. Ziel ist dabei eine zumindest teilweise Automatisierung, die auf der Grundlage von Methoden aus dem Bereich der automatischen Textklassifizierung, des *Machine Learning*, und der Verarbeitung natürlicher Sprache (NLP, *Natural Language Processing*) Vorschläge zur Annotation berechnet, die durch Experten auf ihre Qualität geprüft werden können.
- Um die Metadaten mit Semantik zu hinterlegen, muss das Wissen der jeweiligen Domäne (zumindest teilweise) explizit gemacht und repräsentiert werden. Hierfür eignen sich **Ontologien**, welche die Konzepte einer Domäne erfassen und zueinander in Beziehung setzen. Zur Nutzung von Ontologien im Semantic Web werden sowohl ein erheblicher Aufwand für den konsistenten Aufbau solcher Ontologien notwendig als auch Sprachen zu deren Repräsentation.

Außerdem müssen geeignete Methoden zur Interpretation der semantischen Metadaten entwickelt und diese in Form von effizienten Softwarekomponenten umgesetzt werden. Dabei werden unter anderem Ansätze zur Interpretation der Semantik, zur Realisierung der intelligenten Agenten sowie geeignete Spra-

chen und Protokolle zur Kommunikation zwischen den Agenten benötigt. In der Informatik existieren bereits Ansätze in all diesen Bereichen, die jedoch für die Verwendung im Semantic Web angepasst und erweitert werden müssen. Dabei stellt insbesondere die effiziente Skalierung auf eine so umfangreiche Umgebung wie das (Semantic) Web eine große Herausforderung dar.

Dieser Aufsatz konzentriert sich auf den Aspekt der semantischen Metadaten und ihre Interpretation und gibt dabei auch einen Überblick über die relevanten Ergebnisse der *Semantic Web Activity* des W3C (www.w3c.org).

KONTEXT UND INTERPRETATION

Im Semantic Web und in anderen Umfeldern mit einem Informationsaustausch zwischen unterschiedlichen unabhängigen Kooperationspartnern stellt sich das Problem der Interpretation von Information und Metadaten, die von einer anderen Partei erstellt worden sind. Die Zuordnung von Semantik bei der Interpretation hängt dabei in starkem Maße von deren Kontext ab. Das Wort »Jaguar« oder die Zahl 5, zum Beispiel, können in Isolation betrachtet nicht eindeutig interpretiert werden. Erst durch einen Blick auf den jeweiligen Kontext wird klar, dass sich das Wort »Jaguar« auf das Tier bezieht (und nicht auf das gleichnamige Auto) und dass die Zahl 5 als Preis von 5 Euro zu interpretieren ist.

Menschen können durch ihre kognitiven Fähigkeiten und Erfahrungen diesen Kontext auf vielfältige Weise herleiten, auch wenn er nicht explizit gemacht wird. Zur Interpretation des erwähnten Wortes »Jaguar«, z. B., werden Wörter in dessen Umgebung, das Thema des Textes, in dem das Wort vorkommt, sowie Bilder in dessen Umfeld herangezogen. Der Mensch als Informationsnutzer »sieht«, zum Beispiel, ob eine betrachtete Seite von Autos oder Tieren handelt.

Wesentlich komplizierter stellt sich die Situation dar, wenn die Bedeutung einer Information durch Software interpretiert werden soll. Software kann einen Kontext nur dann zur Interpretation heranziehen, wenn er a) explizit gemacht worden ist und b) geeignet formal beschrieben ist.

Bilaterale Vereinbarungen

Ein nahe liegender Ansatz zum Austausch von Metadaten zwischen zwei unabhängigen Kooperationspartnern, wie z. B. zwischen zwei Organisationen, ist die bilaterale Einigung über die Verwendung eines bestimmten Formats für den Datenaustausch. Eine solche Vereinbarung umfasst neben der Syntax (z. B. XML) die zur Beschreibung zur Verfügung stehenden

Notwendigkeit der Interpretation der Semantik

kontextabhängige Interpretation durch Software

Attribute, d.h. das Vokabular oder Schema zur Datenbeschreibung. Die unterliegende Semantik bleibt in der Regel implizit. Sie wird teilweise in den Programmen codiert, welche die ausgetauschten Daten verarbeiten.

Standardisierung eines Vokabulars

Einen Schritt weiter geht die Vereinbarung bzw. Standardisierung eines Vokabulars/Schemas für eine ganze Nutzergemeinde (*Community*). Ein Beispiel für ein solches Vokabular im Bereich der digitalen Bibliotheken ist der Metadatenstandard Dublin Core (Weibel 1999), der eine Menge von Attributen zur Beschreibung von digitalen Informationsressourcen festlegt. Für einen Datenaustausch zwischen Mitgliedern der Nutzergemeinde kann auf ein in geeigneter Form publiziertes Vokabular zurückgegriffen werden, ohne dass eine detaillierte Absprache über das Format zwischen den Kooperationspartnern notwendig wird. Allerdings ist der Aufwand, sich innerhalb einer Nutzergemeinde auf ein Vokabular zu einigen, nicht zu unterschätzen.

Für die Interpretation von Metadaten kann ein solches Vokabular als gemeinsamer Bezugspunkt für den Kontext verwendet werden. Wenn der Bezug zu dem Vokabular explizit gemacht wird, kann ein Agent oder eine andere Softwarekomponente vergleichbare Felder in der Beschreibung von Ressource erkennen und zueinander in Beziehung setzen. Als Technologie kann hierfür XML und insbesondere das Konzept der XML Namensräume (XML Namespace, [Bray et al. 1999]) verwendet werden. Ein Vokabular kann dabei als Namensraum definiert werden. Durch die systematische Qualifizierung von Namen wird der Bezug zum gemeinsamen Vokabular explizit und auch für eine Softwarekomponente erkennbar gemacht.

Nutzung von Übersetzungsschemata

In der Realität existieren jedoch in einer Nutzergemeinde meist mehrere Vokabulare, die parallel für unterschiedliche Zwecke oder von unterschiedlichen Gruppen verwendet werden. Hierfür gibt es zahlreiche Gründe. Dazu zählen umfangreiche existierende Datenbestände in unterschiedlichen Datenformaten, unterschiedliche Anforderungen von bestimmten Prozessen oder Untergruppen der Nutzergemeinde, etc. Noch schwieriger wird die Situation, wenn Daten zwischen unterschiedlichen Nutzergemeinden ausgetauscht werden sollen.

Für die Bearbeitung von Suchanfragen und die Interpretation von Metadaten in einer solchen inhomogenen Umgebung muss also mit unterschiedlichen Metadaten gleichzeitig umgegangen werden können, um eine konsistente Interpretation aller Daten

zu erreichen. Ein pragmatischer Ansatz hierzu ist die Definition von Übersetzungsschemata zwischen Metadatenformaten, die es erlauben, einen Metadatenatz (automatisch) von einem in ein anderes Metadatenformat zu übersetzen. Im Bereich der digitalen Bibliotheken wurden zu diesem Zweck bereits zahlreiche Übersetzungsschemata, so genannte Crosswalks, entwickelt (siehe z.B. [Baca et al. 1998]).

Ein Agent mit einem Auftrag, der auf der Grundlage eines Metadatenformats A formuliert ist, kann diesen mit Hilfe des entsprechenden Übersetzungsschemas übersetzen, sodass er damit auch eine Kollektion, die gemäß Vokabular B beschrieben ist, durchsuchen kann. Nachteile dieses Ansatzes sind der relativ hohe Aufwand für die Erstellung der Übersetzungsschemata. Außerdem sind nur vorbereitete Kooperationen möglich, d.h. solche, für die bereits ein Übersetzungsschema existiert.

Gemeinsames Datenmodell

Die Formulierung von Übersetzungsschemata und die Kooperation werden erleichtert, wenn den verwendeten Vokabularen ein gemeinsames Datenmodell zugrunde liegt, d.h. ein gemeinsamer Ansatz für die Art, wie die Informationsressourcen beschrieben werden. Der *Resource Description Framework* (RDF, [Manola 2003]), der im Rahmen der Semantic Web Activity des W3C entwickelt wurde und noch weiterentwickelt wird, bietet ein solches Datenmodell zur Beschreibung von Informationsressourcen.

Zentrales Konzept in RDF ist die *Ressource*. Bei einer Ressource kann es sich um digitale und nicht-digitale Objekte handeln, wie z.B. ein Dokument im Web, ein Bild im Museum oder ein Buch in der Bibliothek. Wichtig ist, dass jeder Ressource ein eindeutiger Name zugeordnet ist, ein so genannter *Uniform Resource Identifier* (URI). Dieser ist den im Web verwendeten Adressen (URL, *Uniform Resource Locator*) vergleichbar mit dem Unterschied, dass eine durch eine URI identifizierte Ressource nicht im Web zugreifbar sein muss. Eine URL wird hingegen auch für den Zugriff auf das Objekt verwendet.

RDF stellt ein einfaches, aber mächtiges Datenmodell zur Verfügung. Im Kern werden dabei Aussagen über Ressourcen getroffen, die Eigenschaften von und Beziehungen zwischen Ressourcen beschreiben. Eine Aussage besteht dabei aus einem Subjekt, einem Prädikat und einem Objekt, die alle drei durch eine URI identifiziert werden. Ein einfaches Beispiel ist die Aussage, die als Subjekt auf ein Bild verweist, das Prädikat »wurde gemalt von« besitzt und als Objekt auf den Maler Rembrandt verweist.

Besondere Mächtigkeit erhält man in RDF dadurch,

eindeutiges gemeinsames Vokabular

Konzept der XML-Namespaces

konsistente Interpretation der Daten erforderlich

dass man sowohl RDF-Aussagen als auch Eigenschaften selbst wiederum als Ressourcen betrachten kann, wodurch sich auch Aussagen über Aussagen (*Reification*) und Eigenschaften treffen lassen. So kann man obige Aussage über die Zuordnung eines Bildes zu Rembrandt als Objekt einer anderen RDF-Aussage verwenden und die Aussage treffen, dass ein Kunstexperte (ebenfalls identifiziert durch eine URI) genau diese Zuordnung des Bildes, d. h. die erste Aussage, anzweifelt.

RDF ist nicht auf eine einzige Syntax festgelegt, sondern kann auf verschiedene Weisen repräsentiert werden. Wichtige Syntaxvarianten sind dabei die Formulierung als Graph und die XML-Syntax für RDF (Beckett 2003).

Die Konstrukte von RDF sind dazu geeignet, Aussagen über konkrete Instanzen von Informationsressourcen zu treffen. Zur Definition von Vokabularen wird RDF durch RDF-Schema (Brickley et al. 2003) ergänzt. Mit RDF-Schema lassen sich bestimmte Klassen von Ressourcen, wie z. B. Buch, Herausgeber, Artikel etc. beschreiben. Dabei lassen sich relevante Klassen und Eigenschaften definieren sowie Beziehungen zwischen Klassen, zwischen Eigenschaften sowie zwischen Klassen und Eigenschaften festlegen. Auf diese Weise kann eine Nutzergemeinde ihr eigenes Vokabular aufbauend auf dem Datenmodell von RDF definieren.

Wichtig ist dabei die Möglichkeit von RDF-Schema, Konzepthierarchien aufzubauen, indem Spezialisierungsbeziehungen zwischen Klassen und ihren Subklassen definiert werden. Um auf das Beispiel von weiter oben zurückzukommen, kann z. B. ausgedrückt werden, dass ein Jaguar eine spezielle Art von Säuger ist (Subklassenbeziehung).

Außerdem können mit Hilfe von RDF-Schema Aussagen über Beziehungen zu Konzepten aus anderen Vokabularen getroffen werden. Solche Beziehungen zwischen Konzepten bilden einen ersten Schritt in Richtung der Erfassung von Semantik und können von Agenten im Semantic Web zur Interpretation von Information herangezogen werden bzw., um Informationen aus verschiedenen Kontexten zueinander in Beziehung zu setzen.

Ontologien

»An ontology is a specification of a conceptualization.« (Gruber 1993)

Eine Ontologie repräsentiert eine abstrakte, vereinfachte Sicht, also eine Konzeptualisierung, eines relevanten Ausschnitts der Welt. Sie identifiziert und definiert die relevanten Konzepte einer Domäne und setzt sie zueinander in Beziehung. Damit erfasst eine Ontologie das Wissen über die Domäne, das für die Zuord-

nung von Semantik notwendig ist. Wenn sich Metadaten auf die klar definierten Konzepte einer Ontologie statt auf die Terme eines Vokabular beziehen, werden sie dadurch in einen semantischen Kontext eingebettet, der von den Agenten im Semantic Web zur Interpretation herangezogen werden kann. Auch hier beziehen sich unterschiedliche Nutzergemeinden auf die jeweiligen Ontologien ihrer Domäne. Wichtig für die Interpretation ist in diesem inhomogenen Fall die Nutzung einer gemeinsamen *Top-Level-Ontologie*, die grundlegende domänenunabhängige Konzepte definiert und einen gemeinsamen Bezugspunkt für die Definition domänenspezifischer Konzepte und damit auch für die Interpretation bildet.

Im Bereich der Ontologien befinden sich die Arbeiten im Rahmen der Semantic Web Activity noch in einem relativ frühen Stadium. Mit der *Web Ontology Language* (OWL, [Smith et al. 2002]) liegt eine vorläufige Version einer einfachen Sprache zur Darstellung von Ontologien im Web-Kontext vor, die auf ihrer Vorläufersprache DAML+OIL aufbaut. Sie erlaubt die Definition von Eigenschaften von Beziehungen zwischen Klassen und die Definition von Einschränkungen. Außerdem können Äquivalenzen und Beziehungen zwischen Konzepten unterschiedlicher Ontologien definiert werden. Für die formale Definition der Konzepte selbst bietet OWL keine eigenen Konstrukte an. Dazu muss auf andere Formalismen zurückgegriffen werden. OWL baut auf RDF auf, und es gibt eine gewisse Überlappung zwischen dem, was man mit den beiden Sprachen ausdrücken kann. In diesem Bereich gibt es noch einige offene Forschungsfragen, an denen aktuell gearbeitet wird, wie etwa die konsistente semantische Integration von OWL in RDF.

HERAUSFORDERUNGEN UND ERFOGSAKTOREN DES SEMANTIC WEB

Die Ziele des Semantic Web sind sehr ambitioniert und, da es sich noch in einer frühen Phase der Entwicklung befindet, ist noch nicht klar, ob und in welcher Form sich diese Ziele umsetzen lassen werden. Es gibt eine Reihe von Herausforderungen, von deren Bewältigung es unter anderem abhängt, ob die Vision des Semantic Web erfolgreich umgesetzt werden kann:

- Entwicklung geeigneter Ontologien: Die Entwicklung einer konsistenten Ontologie für eine Domäne stellt eine komplexe Aufgabe dar, an der sowohl Domänenexperten als auch *Knowledge Engineers* mitwirken müssen. Zudem ist eine Domäne nicht statisch, sondern entwickelt sich kontinuierlich weiter, was ebenfalls in der Ontologie reflektiert werden muss.
- Kritische Masse semantisch annotierter Ressourcen: Da die semantische Annotation mehr Aufwand

Top-Level-Ontologie als gemeinsamer Bezugspunkt für die Interpretation

Sprache zur Darstellung von Ontologien im Web

Voraussetzungen für die erfolgreiche Umsetzung der Vision

erfordert als eine Annotation beispielsweise gemäß dem Dublin Core Metadatenstandard, stellt sich die Frage, wie eine hinreichend große Menge an Autoren dazu motiviert werden kann, ihre Ressourcen semantisch zu annotieren. Eine viel versprechende Alternative oder Ergänzung bildet die Entwicklung von Werkzeugen und Diensten zur semi-automatischen Erstellung semantischer Annotationen.

➤ Software zur Interpretation der semantischen Information: Die Entwicklung von Software, welche die semantischen Annotationen effizient und effektiv interpretieren kann, ist eine wichtige Voraussetzung für das Funktionieren des Semantic-Web-Ansatzes.

➤ Entwicklung und Akzeptanz von Anwendungen: Wichtig ist es auch, dass attraktive Anwendungen und Dienste entwickelt werden, die auf dem Semantic-Web-Ansatz beruhen und auch vom Benutzer angenommen werden. Dabei wird neben der Nützlichkeit der Anwendungen auch das Vertrauen eine Rolle spielen, das der Benutzer den Agenten entgegenbringt.

In Anbetracht dieser Herausforderungen ist nicht damit zu rechnen, dass sich das Semantic Web in kurzer Zeit als neues Paradigma für das gesamte Web durchsetzt. Vielmehr ist anzunehmen, dass zunächst viele kleinere Semantic Webs aus gezielten Initiativen entstehen.

Weiter gehende Information zum Semantic Web

Dieser Artikel kann sicher nur einen ersten Einblick in die Idee des Semantic Web geben. Aktuelle und weiter gehende Information zur Semantic Web Activity und den Standards, die in diesem Rahmen entwickelt werden, finden sich auf den Web-Seiten des W3C (www.w3c.org). Einen guten Einblick in die Vision des Semantic Web gibt der bereits erwähnte Aufsatz aus dem Scientific American (Berners-Lee et al. 2001).

LITERATUR

Baca, M.; Busch, J.; Cromwell-Kessler, W.; Gill, T.; Harping, P.: A Crosswalk of Metadata Standards. In: Baca, Murtha (ed.): Metadata: Pathways to Digital Information. Getty Information Institute, 1998, pages 23–33

Beckett, D.: RDF/XML Syntax Specification (Revised) – W3C Working Draft 23 January 2003. URL: <http://www.w3.org/TR/rdf-syntax-grammar>, January 2003

Berners-Lee, T.; Hendler, J.; Lassila, O.: The Semantic Web. In: Scientific American, May (2001)

Bray, T.; Hollander, D.; Layman, A.: Namespaces in XML, W3C Consortium 14–January–1999. URL: <http://www.w3.org/TR/1999/REC-xml-names-19990114/>, January 1999

Brickley, D.; Guha, R. V.: RDF Vocabulary Description Language 1.0: RDF Schema – W3C Working Draft 23 Ja-

nuary 2003. URL: <http://www.w3.org/TR/rdf-schema/>, January 2003

Gruber, T. R.: A translation approach to portable ontologies. In: Knowledge Acquisition, 5(2): 199–220 (1993)

Hendler, J.: Agents and the Semantic Web. In: IEEE Intelligent Systems, 16(2):30–37 March/April (2001)

Manola, F.; Miller, E.: RDF Primer, W3C Working Draft 23 January 2003. URL: <http://www.w3.org/TR/rdf-primer/> January 2003.

Smith, M. K.; McGuinness, D.; Volz, R.; Welty, C.: Web Ontology Language (OWL) Guide Version 1.0, W3C Working Draft 4 November 2002. URL: <http://www.w3.org/TR/owl-guide/>

Weibel, Stuart: The State of the Dublin Core Metadata Initiative – April 1999. In: D-Lib Magazine 5(4) April (1999). URL: <http://www.dlib.org/dlib/april99>

DIE VERFASSERIN

Dr. Claudia Niederée ist Senior Researcher am Fraunhofer Institut für Integrierte Publikations- und Informationssysteme (IPSI) und ist außerdem in der Lehre tätig.

FhG-IPSI, Dolivost. 15, 64293 Darmstadt
niederée@ipsi.fhg.de