

# **Das Semantische Web - Wissensrepräsentation und Schlussfolgerung im Web**

**Jürgen Nicklisch-Franken**

Persist AG –Semantic Web Technology–

[jnf@persistag.com](mailto:jnf@persistag.com)

Phone: 0049-3328-3477-0

Fax: 0049-3328-3477-29

Address: Rheinstraße 7c, D-14513 Teltow, Germany

# **Das Semantische Web - Wissensrepräsentation und Schlussfolgerung im Web**

## *Zusammenfassung*

Bislang war die Struktur des WWW nach dem Abbild von Büchern, bzw. gedruckten Materialien geformt. Die nächste Generation bildet das sog. "Semantische Web", welches auch die spezifischen Möglichkeiten und Fähigkeiten von Computern ausschöpft.

Neben den Zielen des Semantischen Webs werden in diesem Vortrag kurz der heutige Stand der Entwicklung in Beziehung auf die Technologie, die Anwendungen und die Tools beschrieben.

Zu den wichtigen Schritten, um den Übergang vom heutigen Web zum Web der Zukunft zu erreichen, gehört es, dass Metadaten die bisher schon publizierte Information anreichern. Ein wichtiger Teil dieser Metadaten, sind konzeptionelle Modelle, die publiziert werden, und auf die sich ein Autor beziehen kann. Weiterhin gibt es im Semantischen Web neuartige Servertypen, welche die Fähigkeiten haben, Schlussfolgerungen zu ziehen, und die explizit publizierte Information durch nach logischen Regeln gefolgerten Informationen zu erweitern. Dies ermöglicht dann eine Reihe von zusätzlichen Web-Services. Von zentraler Wichtigkeit ist dabei eine neue Arten von Suchmaschinen, die auf spezifische Anfragen genaue Antworten geben können. Die Benutzerschnittstellen sollen einfach und flexibel sein und jedem Benutzer die Möglichkeit geben, nicht nur Konsument sondern auch Autor zu sein.

Zu den wichtigsten kommerziellen industriellen Anwendungsfeldern gehören das Wissensmanagement sowie e-commerce Marktplätze. Schon heute beginnen Unternehmen im Intranet diese neue Form des "Information Brokerings" zu nutzen.

## Ziele des Semantischen Webs

Den Begriff ‚Semantic Web‘ kann man als ‚Netz mit Bedeutung‘ oder in der Manner der deutschen Sprache als ‚Bedeutungsnetz‘ übersetzen. Manche Autoren sind der Meinung, dass ‚Knowledge Web‘ eigentlich ein eingängigerer Ausdruck wäre. Der Begriff stammt offenbar von Tim Berners-Lee, jener Person die das WWW gewissermassen erfunden hat und heute dem W3C, dem World Wide Web Consortium als Direktor vorsteht<sup>1</sup>.

Der Unterschied zum bisherigen Web, welches ja auch Bedeutung transportiert, besteht darin, dass die Daten besser von Computern verarbeitbar sind. Es geht hier also vorrangig nicht um die Bedeutung für Menschen, sondern um Bedeutung für Maschinen. Dem Menschen sollen dann aber die Dienste zugute kommen, die diese Maschinen leisten können. Bisher war die Hilfe die man im Web von Computern erhalten hat eingeschränkt. Suchmaschinen bieten zwar erstaunliche Leistungen was die Schnelligkeit und Zahl der Treffer angeht, aber präzise Anfragen sind nicht möglich. Das Web der Zukunft soll nun Dienste für das Lösen von Problemen anbieten. Ein wichtiger Dienst ist eine Suchmaschine, die präzise Anfragen ermöglicht, z.B.: *‚Welche japanischen Restaurants in Dresden, haben um 22 Uhr noch geöffnet und sind von einer Quelle der ich vertraue, empfohlen worden. Sortiere die Ergebnisse nach dem Preis für Maki-Sushi?‘*.

Ein Semantisches Web, in dem eine Bücherei von Problemlösungsdiensten zur Verfügung stehen, und zwar so, dass ein breites Benutzerspektrum sie nutzen kann, wird auch für die Wirtschaft von einschneidender Bedeutung sein. Das Semantische Netz wird intern als ‚Firmengedächtnis‘ genutzt werden und die technische Grundlage für Wissensmanagement sein. Im Extranet und Intranet wird das Semantische Netz die Plattform für elektronischen Handel in einer neuen Qualität sein.

Jetzt kann man unterschiedlicher Meinung darüber sein, ob die Vision eines Netzes von Diensten (oder anders ausgedrückt Agenten), die auf dem Web tätig sind und komplexe Tätigkeiten ausführen eine ‚Science Fiction Vision‘ ist, oder ob diese Vision sich schon heute zu verwirklichen beginnt. Ich kann hier keine Aussagen über die Zukunft machen, aber ich werde versuchen, soweit mir das möglich ist, eine kompakte Bestandaufnahme zu geben, und es muss dann Aufgabe des Lesers sein, sich sein Urteil zu bilden.

---

<sup>1</sup> In [BL99] wird die Entwicklung des Webs spannend und einfach verständlich beschrieben. Im vorletzten Kapitel ‚Computer und das Web‘ geht es auch um das semantische Web.

## Architektur des Semantischen Webs

**Das Semantische Web braucht eine Grundlage in der Form von Formaten, mit denen man Wissen auf dem Web formalisiert repräsentieren kann. Diese Formate müssen eine Verarbeitung in der Art des Ziehens von Schlussfolgerungen ermöglichen.**

Warum ist dies aber eigentlich notwendig? Auch mit konventioneller Technik werden heute vermehrt Web-Services zur Verfügung gestellt. Man kann heute im Web eine Theaterkarte bestellen, ein Treffen organisieren u.s.w.. Was ist also der Unterschied zum Semantischen Web?

Nehmen wir als Beispiel den Gesundheitsbereich. Hier sind eine Reihe von Parteien beteiligt (Versicherungen, Ärzte, Arbeitgeber, Angestellte), die Informationen zur Verfügung stellen und spezielle Informationsbedürfnisse haben.

*Herr X* ist ein 44 jähriger Feuerwehrmann, der sich bei einem Einsatz *Erfrierungen* am großen Zeh zugefügt hat, weil er versehentlich in einem Kühlraum eingeschlossen war. Er erkundigt sich auf dem Portal seiner Feuerwache nach den *Leistungen*, die er nun in Anspruch nehmen kann. Der Portaldienst des Arbeitgebers ruft einen Dienst des Krankenversicherers auf, um Informationen zu Versicherungsleistungen zu bekommen und lässt sich auch gleich noch von einem Dienst der Berufsgenossenschaft das Vorliegen eines Arbeitsunfalls bestätigen.

Wie können solche Dienste jetzt technisch realisiert werden? Bei der geläufigen Art der Softwareentwicklung wird solche Information anwendungsspezifisch gespeichert (um nicht zu sagen vergraben)<sup>2</sup> und über starre Abfragen ermittelt. Wie eine solche Abfrage bei dem Informationsportal von Krankenversicherer Y auszusehen hat, muss der Dienst der Feuerwehr jetzt speziell wissen. Damit der Feuerwehrmann seine Information bekommt, müssen sich alle Krankenversicherer über einen Standard für Portaldienste einigen, dieser Standard müsste die Abfrage der relevanten Information ermöglichen, der Dienst des Arbeitgebers müsste diesen Standard kennen. Wie schwierig dieses Vorgehen ist, besonders da sich die Angebote und Produkte der Versicherer und aller anderen Beteiligten ständig ändern dürfte klar sein. Die bisherige Art Dienste zur Verfügung zu stellen ist also zu starr und unflexibel um ein Semantisches Web aufzubauen.

Auch bei dem Semantischen Web muss es eine gewisse Vereinheitlichung geben, indem es im Gesundheitsbereich eine Konzeptdefinition für Krankheiten gibt, auf die sich die verschiedenen Parteien beziehen, und *Erfrierungen* ist ein Begriff den

---

<sup>2</sup> Datenbanken bilden hierbei teilweise eine Ausnahme, aber Datenbanken sind nicht mächtig genug um eine strukturierte Information auszudrücken, es fehlt ganz allgemein die Information, welche Bedeutung bestimmte Spalten und Tabellen haben. Aber man kann das Semantische Web in Teilen wie ein universales Wissensbasismanagementsystem im Sinne einer mächtigeren universalen Datenbank ansehen.

man gemeinsam als eine Art von Krankheit kennt, und auch *Erfrierung mit Lokalisation großer Zeh* ist gemeinsam verständlich. Der Dienst der Feuerwehr kann jetzt der Versicherung die Frage stellen, welche Leistungen für *Herrn X*, der der Krankenkasse als Mitglied bekannt bei dieser Krankheit in Anspruch nehmen kann. Die Antwort auf diese Frage bezieht sich wieder auf Begriffe, die gemeinsam verstanden wird. In dieser Form sind natürlich nicht nur Anfragen, sondern auch Aktionen in Form von Meldungen möglich.

Damit die dynamischen Dienste des Semantischen Webs so flexibel sein können, dass der zugrundeliegende Schlussfolgerungsprozess modifizierbar und konfigurierbar ist, benutzt das Semantische Web ein gemeinsames Format mit dem Wissen auf dem Web formalisiert repräsentiert werden kann. Das Wissen des Fachmanns wird in der Form eines Wissensmodells **deklarativ** ausgedrückt. Die Schlussfolgerungsmaschine benutzt dieses Wissensmodell um Anfragen zu beantworten oder Aktionen in die Wege zu leiten.

In den letzten zehn Jahren wurde in der Industrie viel Geld für Modelle ausgegeben, die Wissen über Geschäftsprozesse deklarativ ausdrücken. Im Regelfall wurden diese Modelle dann von den Softwareentwicklern benutzt um das darin enthaltene Wissen in Code zu gießen. Dies ist ein aufwendiges und unflexibles Vorgehen, weil Änderungen in den Modellen zwar relativ einfach gemacht werden können, es aber äußerst aufwendig ist den Programmcode aktuell zu halten. Das Semantische Web kann man in dieser Sicht als eine neue Art Software Plattform ansehen, die auf einer höheren Ebene angesiedelt ist als andere Programmiersprachen, nämlich auf der Wissensebene. Auch die Regeln der Fachlogik werden nicht in prozeduralen Code verwandelt, sondern sind eine Art von Daten und ein Regelsatz wird als ein Webdokument veröffentlicht.

Neben dem deklarativen Vorgehen ist es wichtig, dass die Wissensrepräsentation auf mathematischer **Logik** beruht, und damit ein wohldefiniertes theoretisches Fundament hat. Bei logischen Folgerungen kann die Korrektheit nachgewiesen werden und zwischen verschiedenen logischen Darstellungsformen gibt es erwie-senermaßen funktionierende Mappings. Die Repräsentation als Logik ist an der natürlichen Sprache angelehnt und damit dem menschlichen Denken nah. Gleichzeitig sind die Mittel der Repräsentation auf das wesentliche reduziert. Logik ist damit eine zeitlose Repräsentationsform, und es ist schwer vorstellbar, dass auf dieser Basis IT-Manager gezwungen sind alte Systeme zu verschrotten, weil ‚neue Anforderungen‘ mit ‚neuer Technologie‘ mit den alten Lösungen inkompatibel sind.

Doch Logik alleine reicht nicht aus, für eine Anwendung muss der gewissermaßen leere Rahmen der Logik mit Begriffen gefüllt werden, und dies ist die Aufgabe der **Ontologien**. In Ontologien werden Begriffe kategorisiert und mit Eigenschaften versehen, ganz ähnlich wie man es in einem objektorientierten Modell mit Klassen

macht. Beim Semantischen Web entsprechen diese Begriffe URIs<sup>3</sup>, so dass im Unterschied zu objektorientierten Klassen ein Konzept mit globaler Geltung entsteht.

Damit kann sich ein Web-Dokument auf verschiedene Ontologien beziehen. Wenn ich beispielsweise die Aussage *Arbeitnehmer über 40 Jahren* auf dem Web mache, so ist *Arbeitnehmer* eine URI, die auf eine Ontologie verweist in der dieser Begriff verzeichnet ist. In dieser Ontologie ist ein *Arbeitnehmer* eine Unterklasse von *Person* und ererbt die Eigenschaft *alter*, auf die sich diese Aussage bezieht. Auch der Begriff für eine Eigenschaft wie *alter* ist eine URI.

Ontologien können aufeinander aufbauen und sich aufeinander beziehen. Wenn ein Arbeitgeber sich auf eine Ontologie bezieht, aber dem Begriff des Arbeitnehmers eine Eigenschaft fehlt, die in seinem Betrieb wichtig ist (z.B. *Atembeschwerden*), so kann er eine eigene Ontologie veröffentlichen, die sich auf die andere Ontologie bezieht, und nur für den Arbeitnehmer die fehlende Eigenschaft hinzufügen.

Die dritte Komponente der Wissensrepräsentation auf dem Semantischen Netz betrifft die Verarbeitung des repräsentierten Wissens, also die Schlussfolgerung oder anders gesagt **Inferenz**. Inferenztechniken ermöglichen den Zugriff auf das in der Wissensbasis explizit gespeicherte Wissen, darüber hinaus aber auch auf das implizite Wissen, das aus dem gespeicherten Wissen logisch folgt. Konventionelle Software weist diese Eigenschaft nicht auf. Systeme die bisher Inferenz verwenden sind z.B. Theorembeweiser, logische Programmiersprachen wie Prolog und Expertensysteme.

Systeme die mit Logik und Schlussfolgerung arbeiten, können auch **Beweise** geben, warum sie eine bestimmte Antwort auf eine Anfrage geben. Diese Beweise können dann unter den verschiedenen Diensten ausgetauscht werden und dem Benutzer auf Wunsch präsentiert werden.

Diese Beweise schaffen neben anderen Mittel wie digitalen Signaturen und sicheren Verbindungen die Grundlage für **vertrauenswürdige Dienste**. Informationen dürfen nur an Empfänger gelangen, die auch die Berechtigung dazu besitzen, diese Informationen zu erhalten und gleichzeitig will der Benutzer der Information, die er erhält, vertrauen können. Das bedeutet im Web, dass wenn ich als Benutzer Informationen aus einer bestimmten Quelle erfrage, sicher sein muss, dass dies Information tatsächlich von dieser Quelle kommt. Gleichzeitig gibt es die Möglichkeit auszudrücken, das ich einer Quelle vertrauen will und allen Quellen denen diese Quelle vertraut.

---

<sup>3</sup> URI steht für Uniform Resource Identifier, also zum Beispiel: <http://www.w3.org/Addressing/>. URIs sind eine Verallgemeinerung von URLs (Uniform Resource Locators), die sowohl als Zeiger auf Ressourcen, als auch als Identitäten dienen können.

Das Semantische Web baut auf Technologien der Wissensrepräsentation und Schlussfolgerung auf, die in mehr als 30 Jahren als Teilgebiet des Gebietes Intellektik (künstlichen Intelligenz) entwickelt wurden.

## Das Semantische Web bauen

Um das Semantische Web zu verwirklichen, muss es Standardformate geben, in denen das Wissen abgelegt wird, es muss Softwarewerkzeuge geben, um das Wissen anzulegen und Schlussfolgerungen aus dem Wissen zu ziehen, und es muss schließlich der Aufbau des Wissens selbst erfolgen.

Die Standardformate werden vom W3C als sogenannte Empfehlungen verabschiedet, so wie dies mit HTML und XML geschehen ist. Das Semantische Web benutzt die bestehenden Mechanismen von URIs als Grundlage der Vernetzung und von XML als Grundlage der Syntax.

Das Zentrale Format des Semantischen Webs ist **RDF** (Ressource Description Framework). RDF ist ein Code zur Definition von Informationen im Web, und drückt Dinge, Konzepte, Eigenschaften und Relationen mit Hilfe ihrer URIs aus.

Ontologien werden in **RDF Schema** und **DAML+OIL** ausgedrückt. Während die Empfehlung für RDF vorliegt, ist der Empfehlungsprozess für die Ontologieebene noch in der Entwicklung.

Für die Ebene von Regeln und Beweisen gibt es zwar viele Vorschläge, aber der Prozess der Ausarbeitung einer Empfehlung hat noch nicht einmal begonnen.

Zentrale Tools für das Semantische Web sind **Inferenzmaschinen**, weil erst damit die ganze Mächtigkeit der Konzepte entfesselt werden kann. Die Inferenzmaschinen ermöglichen die Abfrage auch von gefolgertem Wissen und die Implementierung von flexiblen Diensten. Meine Firma arbeitet seit drei Jahren an einem Wissensbasierten Managementsystem für das Semantische Web (Persist<sup>®</sup> SemanticBase), das eine Inferenzmaschine als zentrale Komponente enthält. Wir setzen hier Technologien ein die aus dem Bereich der deduktiven und objektgeduktiven Datenbanken stammen.

Weiterhin wird in den nächsten Jahren ein großer Bedarf dafür entstehen, Web-Ressourcen mit Ontologien zu verbinden und dafür werden **Ontologie- und Metadateneditoren** benötigt. Ein Beispielsystem ist Protegé 2000 ein ‚knowledge-base editing environment‘ das von der Stanford University entwickelt wurde und schon eine längere Geschichte hat, jetzt aber auch zum Editieren von den Formaten des Semantischen Webs benutzt werden kann.

Eine weitere wichtige Klasse von Softwaretools **extrahiert Metadaten** automatisch oder semiautomatisch aus natursprachlichen Texten. Das Worldwide Web

wird sich schrittweise in das Semantische Web transformieren, wenn vermehrt Webseiten mit Metadaten und Ontologien angereichert werden. Diese Tools werden diese Transformation beschleunigen.

Die zur Verfügung stehenden Softwarewerkzeuge für das Semantische Web sind leider nicht so weit entwickelt, wie wir uns das wünschen würden. Aber viele Startup Firmen sind in diesem Bereich tätig und auch etablierte Firmen werden aufmerksam. Ein Teil der zur Verfügung stehenden Tools kommt aus dem akademischen Bereich, in dem die Arbeit an diesem Thema sehr intensiv ist. Ausserdem fördern in den USA die DARPA und in Europa die EU die Entwicklung von Tools für das Semantische Web.

## **Schlussfolgerungen**

Diese Papier ist nicht als umfassender Überblick gedacht, sondern soll eher einen kurzen Einblick in die Arbeit am Semantischen Web geben. Ich hoffe trotzdem, dass ich den Leser überzeugt habe, dass hier ein interessante neue Bewegung aus verschiedenen Strömungen entstanden ist, die einen schon jetzt spürbaren frischen Wind in die IT Landschaft bringt. Die Teile kommen zusammen und es sind jetzt praktische Applikationen, auf denen der Fokus der Entwicklung liegt.

## **Links**

World Wide Web Consortium (W3C) -> [www.w3.org/](http://www.w3.org/)

Aktivitäten des semantischen Netz -> [www.w3.org/2001/sw/](http://www.w3.org/2001/sw/)

Semantic Web Business Special Interest Group-> [business.semanticweb.org](http://business.semanticweb.org)

DARPA-Projekt über agentengestützte Computertechnik -> <http://www.daml.org/>

## **Referenzen**

[BL99] Tim Berners-Lee, Der Web-Report, München 99

[BI92] Wolfgang Bibel, Deduktion, München 92

[BI93] Wolfgang Bibel, Wissensrepräsentation und Inferenz, Braunschweig 93

[SO00] John F. Sowa, Knowledge Representation, Brooks/Cole 00