



## Semantisches Web

Tassilo Pellegrini, Andreas Blumauer (Hrsg.), *Semantic Web. Wege zur vernetzten Wissensgesellschaft*, Springer Verlag, Heidelberg, Berlin 2006. ISBN 3-540-29324-8

Einige Vorbemerkungen sollen hier eine Begründung liefern, wieso dieser Band auf einer Website für Romanisten besprochen wird. Allein schon weil das Semantische Web sich einen Begriff aus der Sprachwissenschaft ausgeliehen hat, ist es sicher auch für Romanisten und überhaupt auch für Philologen interessant, die Verwendung dieses Begriffes in einem neuen Zweig der Internet-Technik zu verfolgen

Die Suche nach dem Begriff > [Romanistik](#) mit einer Suchmaschine offenbart das bereits hinreichend bekannte Problem einer langen nach einem Suchalgorithmus von einer Maschine mit wenig romanistischem Sachverstand geordneten Liste. Aber auch Linklisten wie die auf der Website > [Romanistik im Internet](#) sind eher eine Momentaufnahme, die nicht immer die Interessen oder die Suchbedürfnisse der Besucher berücksichtigen kann. In diesem Sinn ist eine traditionelle Bibliothek mit fachmännischer Verschlagwortung der heutigen Internet-Ordnung immer noch überlegen.

Das Problem der Ordnung von Informationen gibt es nicht erst seitdem das Internet entstanden ist. Die vielfältigen Vernetzungen jeder Art haben eine Orientierung im Datendickicht des Internets nicht unbedingt erleichtert, sondern die Koexistenz von einleuchtenden, weil vom Sachzusammenhang gebotenen Verlinkungen und solchen, die willkürlich aufgrund bestimmter Individueller Interessen gesetzt wurden, erschweren die Auswahl und die Bewertung von Informationen. Die Organisation der Informationen auf Websites hat sich ihre eigenen Gesetzmäßigkeiten geschaffen, oder zumindest wird dies von schnell hergestellten Websites zumindest suggeriert. Es gibt genügend Beispiele aus der Arbeit mit Internet-Seiten, die Wissenschaftler, die eher mit traditionellen Mitteln arbeiten, mit Recht in Erstaunen versetzen. Das Zitieren aus der Online-Enzyklopädie

> [Wikipedia](#) gehört zu dieser Art von Beispielen, die diesem Projekt einen wissenschaftlichen Anspruch verleihen, der in keiner Weise zu rechtfertigen ist. Schon fangen Studenten an, die Inhalte der Online-Enzyklopädie als Zitatenschatz für Seminararbeiten zu nutzen. Ein Gewährsmann für die Inhalte der Artikel wird nicht mehr benötigt und der inhaltlichen Willkür wird Tür und Tor geöffnet, weil dem Online-Projekt, an dem jedermann mitschreiben darf, eine Autorität zuerkannt wird, die die Kontrolle einer anonymen Gemeinschaft der Gewährleistung durch einen individuellen Autor vorzuziehen scheint. Das Vertrauen, das in dieses Online-Projekt gesetzt wird, entspricht dem Vertrauen, daß meist aus Bequemlichkeit in die Ergebnislisten der Suchmaschinen gesetzt wird.

Die Suche nach > [Semantic Web](#) mit Hilfe von Google führt zu rund 90 Mio. Webseiten in einer bestimmten Reihenfolge, nicht weil sie die besten Informationen zu diesem Thema enthalten, sondern lediglich, weil der Suchalgorithmus von Google die gefundenen Seiten in eine bestimmte Reihenfolge stellt, die mit der Qualität oder ihrem Informationsgehalt kaum etwas zu tun haben. Man verweilt oft aus Zeitgründen bei den ersten aufgerufenen Seiten und verleiht ihnen so möglicherweise eine Bedeutung, die eine genaue Durchsicht vieler weiterer Seiten nicht rechtfertigen würde. Mehrdeutige Suchbegriffe werden dabei

nicht erkannt und tragen kaum dazu bei, das Suchergebnis qualitativ zu verbessern. Immerhin geben Suchausdrücke, also die Kombination mehrerer Wörter häufigen Benutzern dann doch zu erkennen, daß Google letztendlich eben doch nur eine indexbasierte Suchmaschine ist. Google nutzt heute u.a. die Zahl der Links, die auf eine Seite zeigen, um deren „Qualität“ zu bewerten. Die Höhe des „Pageranks“, der von Suchalgorithmus ermittelt wird, entscheidet über die Platzierung des Website im Suchergebnis. Das Verfahren ist wissenschaftlich gesehen untauglich, trotzdem verwenden Betreiber von Websites viel Energie darauf dieses Spiel mitzuspielen.

Schon in den Anfangsjahren des Internets Mitte der 90er Jahre stellte sich immer dringlicher die Frage nach einer Orientierung im Internet, die sich eines Tages von mehr oder weniger automatisch erstellten Listen lösen kann und Methoden einführt, die den Suchergebnissen eine gewisse Qualität verleihen. 1998 trug Tim Berners-Lee einen neuen Ansatz vor, der als Semantic Web<sup>1</sup> bekannt wurde. Sein Gedanke zielte auf eine Standardisierung der Verfahren, wie Informationen beschrieben werden sollten. Gemeint waren Zuordnungen, die es ihrerseits "intelligente Agenten" erleichtern würden, Informationen aus unterschiedlichen Quellen miteinander zu verbinden. Ähnlich wie in der Sprachentwicklung werden bei diesem Schritt Informationen im Internet Bedeutungen verliehen oder zugeordnet, wodurch das Netz um eine semantische Dimension erweitert wird. Um die Zusammenhänge zwischen diesen Bezügen erkennbar zu machen, sprach schon Berners-Lee davon, Ontologien einzuführen, die in größeren Zusammenhängen Wissensbereiche beschreiben und die Bezüge zwischen Objekten aufdecken sollen. Dieser der Philosophie entlehnte Begriff dient dazu, einen Dialog zwischen einem Anwender und der Maschine zu beschreiben und folglich auch zu steuern. Im Kern bezeichnet eine Ontologie viele Verhältnisse zwischen Informationseinheiten (Entitäten) und den Zusammenhängen, worauf sie sich beziehen, den Relationen. Fragt der Nutzer nach „Bank“ können Suchsysteme, die mit Ontologien arbeiten, denen also definierte Wissensbereiche bekannt sind, zu einem solchen mehrdeutigen Begriff passende Ergebnisse oder zumindest weitere Suchbegriffe vorlegen. Ob allerdings Suchmaschinen eines Tages Fragen nach bestimmten Sachverhalten beantworten können, sei dahingestellt. Das semantische Netz ist der Versuch, die vielen bisher meist auf der Grundlage von HTML entwickelten Websites, die eigentlich kaum mehr als durch Links untereinander verbunden sind, in ein wissensbasiertes Netz zu überführen.

Mit den Metatags, wie z.B. den „Keywords“, die auf HTML-Seiten eingetragen werden können, und die den Suchmaschinen eine Ordnungshilfe gewähren sollten, wurde wohl viel Mißbrauch getrieben, so daß Google diese Hinweise nicht mehr berücksichtigt. Theoretisch könnten semantisch orientierte Systeme einer ähnlichen Gefahr ausgesetzt werden, die möglicherweise durch webbasierte Systeme, die vom Betreiber der Seite nicht zu beeinflussen sind, vermindert oder vermieden werden. Ob aber dann nicht neue Probleme auftauchen, wenn der Bedeutungsgehalt der Seite von Programmen definiert werden soll, ist noch gar nicht abzusehen. Ein falsch getaggeter Text könnte dann das gleiche Schicksal wie ein in einer Bibliothek verstelltes Buch erleiden, das für die Ausleihe nicht mehr zur Verfügung steht.

---

<sup>1</sup> Tim Berners-Lee, James Hendler and Ora Lassila, A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities > <http://www.scientificamerican.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21&catID=2> .

Der gerade erschienene Band *Semantic Web. Wege zur vernetzten Wissensgesellschaft*, (Inhaltsverzeichnis: > [www.semantic-web.at/springer/](http://www.semantic-web.at/springer/)), dessen Herausgeber > [Tassilo Pellegrini](#) von der > [Semantic Webschool](#), ein Zentrum für Wissenstransfer in Wien, kommen, bietet einen interessanten Einblick in das Konzept und das Programm dessen, was der Text auf dem Buchrücken als die „nächste Generation des Internets“ bezeichnet. Der Untertitel des Bandes trifft nicht genau das eigentliche Thema dieses Bandes. Es geht nicht nur um bloße Vernetzungen, sondern um den Beitrag semantischer Technologien, d.h. Verfahren und Methoden, die sich explizit auf die Erstellung und Maschinenlesbarkeit von Bedeutungsträgern konzentrieren. Es geht um Methoden zur qualitativen Orientierung im Internet.

Die neuesten Trends im Internet, die mit dem Schlagwort Web 2.0 gekennzeichnet werden,<sup>2</sup> drücken sich auch durch die sprunghafte Zunahme von Blogs aus. Eines der Kennzeichen von Web 2.0 ist eine besonders große Zunahme unstrukturierter Daten, die durch das Tagging der User nur behelfsweise geordnet werden können. Hier setzen die Überlegungen ein, die mit dem Semantic Web in Verbindung stehen, und die sich auf eine Reduzierung, Verdichtung und Strukturierung der Daten beziehen.

Die Linguistik lehrt, daß Semantik „sich mit der Analyse und Beschreibung der sogen. ‚wörtlichen‘ Bedeutung von sprachlichen Ausdrücken beschäftigt.“<sup>3</sup> Die Semantik wird von der Morphologie (Flexion- und Wortbildungslehre) wie auch von Syntax (Anordnung von Zeichen) abgegrenzt, die ihrerseits ihren Teil zur Sinnkonstituierung beitragen. Die Semantik konzentriert sich nach V. Nyckees auf die Bedeutungsaspekte, die nicht mit anderen Bedeutungsträgern Ähnlichkeiten oder Beziehungen teilen, sondern die sich auf das Wissen einer Sprachgemeinschaft beziehen.<sup>4</sup> In diesem Sinn bildet die Sprache ein ähnliches System von Konventionen, wie das semantische Web diese im Internet einführen will. In diesem Sinn gibt es sicher eine gewisse Berechtigung, diesen Begriff zu verwenden. Vielleicht stutzt der Leser dann doch, wenn er auf dem Buchrücken des hier zu besprechenden Buches liest: „Semantik ist (...) ein wesentliches Element der Transformation von Information in Wissen, sei es um eine effizientere Maschine-Maschine-Kommunikation zu ermöglichen oder um Geschäftsprozeß-Management, Wissensmanagement und innerbetriebliche Kooperation durch Modellierung zu verbessern.“ Es ist keinesfalls die Semantik die das Entstehen von Wissen bewirkt, allenfalls beschreibt sie Prozesse, wie Bedeutungen entstehen. Ein Buchrücken ist geduldig, und Raphael Capurro hat den Begriff der Semantik in seinem sehr lesenswerten Nachwort „Hermeneutik revisited“ zu diesem Band in einen richtigen Zusammenhang gerückt. Er erklärt auf einprägsame Weise den Zusammenhang zwischen der Hermeneutik und der Semantik.<sup>5</sup> Nach dem griechischen Götterboten Hermes bedeutet Hermeneutik die Verkündung, Auslegung oder Erklärung. Capurro nennt auch Wittgenstein und dessen „Sprachspiele“, die nach seiner Theorie dazu geeignet sind, zwischen verschiedenen Gebilden Bezüge herzustellen und

<sup>2</sup> Tim O'Reilly, *What Is Web 2.0? Design Patterns and Business Models for the Next Generation of Software* > <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>

<sup>3</sup> H. Bußmann, *Lexikon der Sprachwissenschaft*, Stuttgart 2/1990, S. 672.

<sup>4</sup> Vgl. Nyckees, Vincent, *La sémantique*, Paris 1998, S. 11.

<sup>5</sup> Zur Hermeneutik, cf. Vogt, Jochen, *Einladung zur Literaturwissenschaft*, 3. Aufl., Stuttgart 2002, Kapitel 3: > <http://www.uni-essen.de/einladung/Vorlesungen/hermeneutik/main.html>; H.-G. Gadamer, *Wahrheit und Methode*, Tübingen 1975; id., *Semantik und Hermeneutik*, in: id., *Hermeneutik II. Wahrheit und Methode. Ergänzungen*, Tübingen 1986, S. 174-183.

sie so verständlich zu machen. Gerade die Abstraktionsfähigkeit des Menschen ermöglicht es, Bezeichnungen für Maschinen und Computer verstehbar zu machen, sie also in digitale Kombinationen von 1 und 0 zu übersetzen. Diese Strukturen behalten aber nur ihren Sinn, weil die menschliche Welt trotz der immer weiter ausufernden Datenmengen dem Internet erst einen Sinn verleiht. Capurro weist selbst daraufhin, daß das Verstehen nicht auf eine Eigenart des jeweiligen Subjekts ist, sondern eine „Seinsweise“ (vgl. R. Capurro, *Hermeneutik der Fachinformation*, Freiburg/München 1986, S. 11) bezeichnet, der der Mensch sich nicht entziehen kann. Auf diese Weise entsteht die bereits angedeutete Verwendung der Ontologie im Rahmen des Semantischen Webs.

Capurro vertritt mit Recht der Auffassung, daß die digitale Welt gegenüber der Lebenswelt keinerlei Vorrang besitzt. Allerdings sollte auch die „Kraft des Konkretion“ (S. 531) der Abstraktion der digitalen Welt entgegengestellt werden. Das Semantische Web wird sich nur im Rahmen menschlichen Handelns entwickeln, wenn dieses jeder Abstraktion begegnen kann. Damit nennt Capurro eine ethische Dimension des Semantischen Webs, das er mit einer Art neuen Hermeneutik verknüpft, die die Bewahrung und Verteilung digitaler Inhalte betrifft. Auf diese Weise erklärt er das Semantische Web als ein „weltpolitisches Projekt“ (S. 532), das nicht Technikern oder Politikern überlassen werden darf.

An diesem Band sind 57 Autoren aus 35 Institutionen beteiligt. Es geht darum, Technologien für das Internet zu entwickeln, die sich auf Verfahren stützen, die semantische Bedeutungsträger erkennen und auswerten können. Daten wie E-Mails können heute schon mehr schlecht als recht durch Programme sortiert werden, die z. B. als Spam-Filter, unliebsame E-Mails ausfiltern, oder es gibt auch Programme, die z.B. E-Mails an den richtigen Arbeitsplatz verteilen.

In ihrem Vorwort stellen die Herausgeber drei Trends vor: Es ist eine steigende Nachfrage nach dynamischen Produkt- und Dienstleistungskonfigurationen zu erkennen. Damit ist eine Dynamisierung der Arbeitswelt verbunden, die organisationsübergreifende Strukturen erfordert. Das Anwachsen der Informationsbestände erfordert ständig intelligenter Dialog- und Suchwerkzeugen, um die verlangten Dienstleistungen erbringen zu können. Diese Art der Kreisbewegung der steigende Nachfrage, Dynamisierung und Bereitstellung intelligenter Suchsysteme bestimmt folglich den Aufbau dieses Bandes.

Der erste Teil dieses Bandes enthält eine Reihe von einführenden Artikeln, in denen zuerst Begriffe (u.a. Semantic Web und semantische Technologien, A. Blumauer, T. Pellegrini) und Standards (Standards für das Semantic Web, K. Birkenbühl) erläutert werden. M. Weber und K. Fröschl untersuchen „Das Semantic Web als Innovation in der ökonomischen Koordination“ und geben interessante Hinweise auf die „Innovationspotenziale semantischer Technologien“ (S. 105 ff). Im zweiten Teil geht es um die Anwender des Semantic Web. L. Sauer mann untersucht den „Semantic Desktop – Der Arbeitsplatz der Zukunft“. Seine systematische Darstellung mit vielen praktischen Beispielen ist auch gut als Einstiegsartikel in das Thema dieses ganzen Bandes geeignet. „Knowledge Visualization“ ist das Stichwort, mit dem Remo Burkhard die Frage nach der „nächsten Herausforderung für Semantic Webforschende“ stellt. Er erinnert daran, daß eine allgemeine Visualisierungswissenschaft mit einem entsprechenden Theorierahmen noch immer fehlt. Die Beispiele, die er in seinem Beitrag nennt, zeigen die Ansätze mit denen Verbindungen wie die zwischen der Bildwissenschaft und anderen Disziplinen erarbeitet werden. Der dritte Teil untersucht die Bedingungsfaktoren für das Semantische Web unter dem Aspekt des

Wissensmanagement. Schmitz et. al. erläutern einen solchen Ansatz auf einer Peer-to Peer-Basis, während Hannes Werthner und Michael Borovicka die praktischen Zusammenhänge zwischen E-commerce und Semantic Web am Beispiel von „Harmonise“, einem EU-Projekt aus der Tourismus-Branche, das mit intelligenten „Networkings“ arbeitet, vorstellen. Im vierten Teil geht es wieder um die Theorie und die technischen Systeme, aber auch um praktische Ansätze, so wie der Beitrag von Andreas Koller, der nachzuweisen versucht, daß eine strukturierte Ablage von Content in Content Management Systemen, eine Voraussetzung für das Semantische Web ist. Mit vielen Beispielen gelingt es Koller die unterschiedlichen Bedingungen für strukturierten Inhalt dazulegen; zugleich wird aber auch die Komplexität des Semantischen Webs deutlich. Computerlinguisten werden sich für den von Michael Granitzer verfaßten Aufsatz im vierten Teil dieses Bandes interessieren. Er stellt statistische Verfahren vor, die dringend benötigt werden, ist doch der digitale Datenbestand bereits auf das 37.000-fache der Library of Congress angewachsen, die ca. 17 Millionen Bände besitzt. Merkmalsanalysen, Lemmatisierung, Parsing und Kollokationen gehören zu den Stichwörtern, die Granitzer untersucht, um die Bedingungen der Textanalyse zu erläutern. Gerhard Budin erläutert die „Kommunikation in Netzwerken“ Marc Ehrig und Rudi Studer stellen die „Wissensvernetzung durch Ontologien“ vor. Ihnen geht es um die Interoperabilität zwischen unterschiedlichen Webservices, die nur, wie sie es nennen, durch eine semantische Integration der Ontologien möglich wird.

Der mit dem semantischen Web verbundene Anspruch wird in der Forschung gerade erst formuliert, noch steht die Realisierung der entsprechenden Anwendungen ganz am Anfang, aber die Hoffnungen, die in dieses Projekt gesetzt werden, lassen die Dimension dieses Projekts erahnen: „Das semantische Web ermöglicht neben erheblichen Verbesserungen der Usability, einen höheren Gebrauchswert verfügbarer Informationsbestände und effizientere Wissensströme.“<sup>6</sup>

Dieser Band bietet eine gelungene Einführung in das Thema, weil sich seine Autoren an ganz praktischen Beispielen orientieren, viele verschiedene theoretische Ansätze vorstellen und insgesamt die Perspektiven für das Semantische Web in einer spannenden Weise vermitteln. Außerdem ist es den Herausgebern gelungen, durch die Auswahl der Beiträge, die Bedeutung der beteiligten Disziplinen in einleuchtender Weise zu vermitteln.

Heiner Wittmann

*Bibliographische Angaben:*

Bußmann, Hadumod, *Lexikon der Sprachwissenschaft*, Stuttgart 2/1990.

Gadamer, Hans-Georg, *Wahrheit und Methode*, Tübingen 1975

-- *Hermeneutik II. Wahrheit und Methode*. Ergänzungen., Tübingen 1986.

Nyckees, Vincent, *La sémantique*, Paris 1998.

O'Reilly, Tim, What Is Web 2.0? Design Patterns and Business Models for the Next Generation of Software > <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>

Vogt, Jochen, *Einladung zur Literaturwissenschaft*, 3. Aufl. Stuttgart 2002. > <http://www.uni-essen.de/einladung/>

Berners-Lee, Tim, Hender, James, Lassila, Ora, A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities, > <http://www.scientificamerican.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21&catID=2>

<sup>6</sup> Andreas Blumauer, Tassilo Pellegrini, > [Semantisches Web – schon wieder eine Patentlösung für die Wissensgesellschaft?](http://www.semantic-web.at/36.20.20.article.kontext.semantisches-web-schon-wieder-eine-patentloesung-fuer-die-wissensgesellschaft.htm) > <http://www.semantic-web.at/36.20.20.article.kontext.semantisches-web-schon-wieder-eine-patentloesung-fuer-die-wissensgesellschaft.htm>