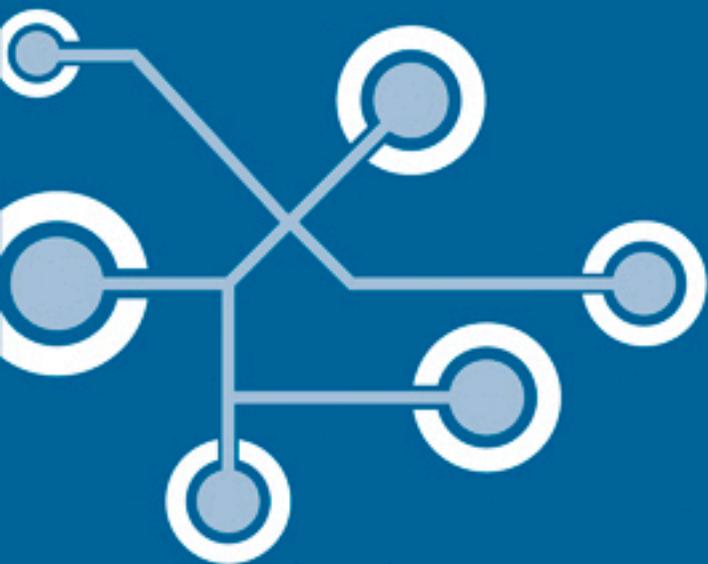


entwickler.press

Semantic Web

Matthias Geisler



schnell + kompakt

Semantic Web

Matthias Geisler

Semantic Web

schnell+kompakt

entwickler.press

Matthias Geisler
Semantic Web
schnell+kompakt
ISBN: 978-3-86802-028-1

© 2009 entwickler.press
ein Imprint der Software & Support Verlag GmbH

<http://www.entwickler-press.de>
<http://www.software-support.biz>

Ihr Kontakt zum Verlag und Lektorat: lektorat@entwickler-press.de

Bibliografische Information Der Deutschen Bibliothek
Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.ddb.de> abrufbar.

Lektorat: Sebastian Burkart, sburkart@entwickler-press.de
Korrektur: Frauke Pesch
Satz: Sebastian Burkart
Umschlaggestaltung: Maria Rudi
Belichtung, Druck und Bindung: M.P. Media-Print Informationstechnologie GmbH, Paderborn.
Alle Rechte, auch für Übersetzungen, sind vorbehalten. Reproduktion jeglicher Art (Fotokopie, Nachdruck, Mikrofilm, Erfassung auf elektronischen Datenträgern oder andere Verfahren) nur mit schriftlicher Genehmigung des Verlags. Jegliche Haftung für die Richtigkeit des gesamten Werks, kann, trotz sorgfältiger Prüfung durch Autor und Verlag, nicht übernommen werden. Die im Buch genannten Produkte, Warenzeichen und Firmennamen sind in der Regel durch deren Inhaber geschützt.

Inhaltsverzeichnis

Vorwort	7
Danksagungen	9
Widmung	9
Der Autor	9
Quelltext-Download der Beispiele	10
Einleitung	11
Kapitel 1: Grundlagen	15
1.1 Begriffsdefinition	15
1.2 XML	17
1.3 RDF	27
1.4 Web-Ontologiesprachen	40
Kapitel 2: Semantic-Web-Praxis	43
2.1 Mikroformate	43
2.2 Dublin Core	62
2.3 Friend of a Friend	69
2.4 RDFa	76
2.5 Linked Data	83
Kapitel 3: Erweiterte Konzepte	95
3.1 Einfachere Ontologien mit RDF Schema	96
3.2 Die Web-Ontologie-Sprache OWL	102
3.3 Die RDF-Anfragesprache SPARQL	112
Kapitel 4: Anhang	121
4.1 Die eigene Linked-Data-Webseite mit Paget	121
4.2 RDF-Datenspeicher und SPARQL-Client mit Drupal	131
Stichwortverzeichnis	143

Vorwort

Das Internet ist gerade dabei, den nächsten Schritt in seiner Evolution zu gehen, zum Web 3.0. Das Web 1.0 war ein einfaches, nur zu konsumierendes Web, der Nutzer konnte nur HTML anschauen, lesen und selbst interpretieren. Web 2.0 erlaubt es dem Nutzer, selbst zu agieren, in dem er bloggt, chattet, bewertet, Communities betritt etc. Das Web 3.0 führt uns über Linked Data zum Semantic Web. Hier wird nicht mehr nur der Mensch die im Internet vorhandenen Informationen interpretieren und verstehen können, sondern in zunehmendem Maße die Computer und Programme selbst.

Dies eröffnet die Möglichkeit zu vollkommen neuen Angeboten und Dienstleistungen. Die angebotenen Informationen und Leistungen werden dem Nutzer immer zielgenauer und personalisierter zur Verfügung stehen, weil die Computer auf der einen Seite die Anfrage, d. h. die in der Anfrage enthaltene Semantik (Bedeutung) verstehen werden und auf der anderen Seite die vorhandenen Informationen im Internet auch semantisch erschlossen haben, wodurch ein zielgenaues Matching erreicht wird.

Nehmen wir ein einfaches Beispiel: Sie wollen z. B. wissen, wie die Ehefrau von Helmut Schmidt heißt, so können Sie googeln, aus den 730 000 Treffern eine Biografie suchen und den entsprechenden Abschnitt über sein Privatleben lesen. Eine Möglichkeit!

Sie können aber auch in der DBpedia auf die Seite <http://dbpedia.org/sparql> gehen, eine gezielte Anfrage stellen und genau die EINE gesuchte Antwort finden. Zugegeben, die SPARQL-Anfrage

```
SELECT ?name
WHERE {
  dbpedia:Helmut_Schmidt dbpprop:spouse ?name
}
```

ist nicht gerade jedermanns geläufigste Sprache, aber mit Sicherheit werden in naher Zukunft natürlichsprachliche Zugänge existieren, und Sie werden einfach eintippen „Ehefrau von Helmut Schmidt“ und genau die eine Antwort erhalten, die Sie suchen: Hannelore „Loki“ Glaser.

Im aktuellen Forschungsprojekt der Bundesregierung THESEUS¹ werden gerade die Techniken und Systeme entwickelt, die es Ihnen dann auch noch erlauben, z. B. die Tagesschau-Szenen zu finden, in denen Hannelore Schmidt zusehen ist, die Abschnitte von Bibliografien anzuzeigen, in denen über sie berichtet wird, oder eine Radiosendung über sie zu hören.

Damit dies möglich wird, werden die im Internet vorhandenen Informationen zunehmend mit semantischen, maschinenlesbaren Informationen angereichert (annotiert). Dies geschieht nicht nur durch Menschen, wie in der DBpedia, sondern auch mit wachsender Qualität automatisiert, wie dies z. B. OpenCalais² zeigt.

Dieses Buch führt Sie in die grundlegenden Standards und Sprachen des Semantic Webs ein, die diesen nächsten evolutionären Schritt möglich machen, und zeigt Ihnen an einfachen Beispielen, wie Sie diese selbst verwenden und anwenden können.

Viel Spaß bei der Lektüre!

Dr. Christoph Schmidt

1 <http://theseus-programm.de>

2 <http://www.opencalais.com/>

Danksagungen

Danken möchte ich meiner Freundin Anne für Ihre Geduld und Unterstützung – deren Grenzen auszuloten eine Herausforderung wäre.

Für das wissende und faire Fachlektorat, das spitzfindige Vorwort und den unkomplizierten und offenen Kontakt bedanke ich mich sehr herzlich bei Dr. Christoph Schmidt.

Bei dem Team von *entwickler.press*, allen voran Sonja Waldschuk und Sebastian Burkart, möchte ich mich für die professionelle Unterstützung und die stets unkomplizierte Zusammenarbeit bedanken. Für den Kontakt zum Team von *entwickler.press* und die Ermutigung zu diesem Buchprojekt gilt mein Dank Raju Bitter.

Widmung

Dies Buch, klein aber fein, widme ich meiner Freundin Anne und meinen Eltern – für Eure Liebe.

Der Autor

Matthias Geisler, geboren 1976, ist Dipl. Wirtschaftsinformatiker (BA) und lebt in Donauwörth. Er arbeitet als Senior Consultant, Projekt- und Teamleiter mit den Schwerpunkten moderne Webtechnologien, Open-Source-Software/Open-Web-Standards und Web/Enterprise 2.0 in München.

Sie erreichen ihn per E-Mail unter *matthias@semanticdreamer.com* oder über sein Weblog *http://semanticdreamer.com*.

Quelltext-Download der Beispiele

Begleitend zum Buch stellt der Autor über sein Weblog unter <http://semanticdreamer.com/semantic-web-schnell-und-kompakt/> die Beispiele aus diesem Buch als Quelltext zum Download sowie – wo sinnvoll möglich – auch als Onlinedemo bereit.

Einleitung

Das World Wide Web (WWW), oftmals und fälschlicherweise synonym mit dem Internet gleichgesetzt, ist heute weder aus dem privaten noch aus dem beruflichen Alltagsleben eines Menschen in den Industriestaaten wegzudenken. Und das gilt beileibe nicht mehr nur für den technikvisierten und -interessierten „Nerd“ oder den modernen Wissensarbeiter. Zudem wird das Internet, und mit ihm das WWW, in schnell wachsendem Ausmaß insbesondere auf Grund seiner standardbasierten, fehlertoleranten und einfach zugänglichen Auslegung als globales Medium für den Informationsaustausch auch in den Nicht-Industriestaaten zu einem immer wichtigeren und vor allem immer stärker verbreiteten und zugänglichen Medium.

Gerade die in unseren Breiten schnell wachsende Anzahl von Breitbandinternetzugängen verändert darüber hinaus unser tägliches Nutzungsverhalten des WWW: Wir suchen und suchen und finden etwas (vielleicht das Gesuchte), wir kommunizieren miteinander, wir informieren uns, wir kaufen ein, wir planen und verabreden uns, seit dem Web 2.0 (oder auch „Read/Write Web“) publizieren wir nun auch vermehrt selbst unsere Ideen, Ansichten und Meinungen – und das alles, versteht sich, jederzeit und überall!

Aber diese Aktivitäten müssen im heutigen WWW zum Großteil noch sequenziell als einzeln und nacheinander ausgeführte Aktionen stattfinden. Das jeweilige Ergebnis bedingt zwar unter Umständen eine weitere Aktion bzw. dient als ihre Eingabe, diese muss aber manuell bzw. mittels „Copy-and-Paste“ übertragen werden. Das folgende Beispiel illustriert unseren heutigen Umgang mit dem Web 2.0.

BEISPIEL

Der perfekte Abend mit unserem neuen Schwarm Claudia soll wie folgt aussehen: „Ich hole Claudia bei Ihr zu Hause ab und führe sie fein zum Essen aus – sie isst gerne italienisch! Danach gehen wir in die Spätvorstellung des neuen James-Bond-Kinofilms, und zwar in ihr Lieblingskino, von dem sie mir erzählt hat (angeblich ganz bei ihr in der Nähe). Und dann... sehen wir weiter.“

Da wir natürlich immer online sind, wird dieser Plan wie folgt im WWW recherchiert:

1. **Orientieren: Wo wohnt Claudia?**

Die Adresse aus Ihrer letzten E-Mail per „Copy-and-Paste“ nach Google Maps kopieren und nachschlagen.

2. **Das Kino finden, bzw. „googeln“**

Verschiedene Suchbegriffe erfolglos bei Google ausprobieren, schließlich die Webseite mit dem gesammelten Kinoprogramm unserer Stadt durchscrollen und grob die Adressen im Kopf lokalisieren – das muss es sein! Die Anschrift noch mal schnell und wieder per „Copy-and-Paste“ mit Google Maps verifizieren, stimmt!

3. **Die Kinokarten reservieren**

Auf der Webseite des Kinos die Telefonnummer herausuchen und abschreiben – natürlich gibt es keine Möglichkeit zur Online-Kartenreservierung. Telefonisch zwei Karten reservieren.

4. **Ein gutes italienisches Restaurant in der Nähe finden**

Eine Google-Suche, diesmal mit unserer Stadt und „italienisch+restaurant“ liefert zwar tatsächlich Treffer, und auf einer kleinen Karte finde ich auch bunte Markierungen in der Nähe des Kinos – nur welches ist nun „gut“? Ich erinnere mich an den Tipp eines Freundes und suche auf einer Web-

2.0-Community-Plattform nach unserer Stadt. Über den Filter „Restaurants“ und die integrierte Google-Maps-Karte finde ich dort tatsächlich ein von anderen Nutzern der Plattform mit 5 Sternen bewertetes italienisches Restaurant ganz in der Nähe, außerdem den Link zur Homepage des Restaurants (dabei fällt mir auf, dass Google das Restaurant vorhin gar nicht gefunden hatte!).

5. **Einen Tisch für zwei reservieren**

Auf der Webseite des italienischen Restaurants die Telefonnummer finden und abschreiben – die Nummer steht auf dem Foto von Venedig im Kopfbereich der Seite (auch hier gibt es natürlich keine Möglichkeit zur Online-Tischreservierung) – und telefonisch einen Tisch für zwei reservieren.

Das beschriebene Beispiel kennen wir in seinem Kern so oder so ähnlich wohl alle. Wir nutzen das WWW und seine diversen Angebote zwar täglich, aber dabei beschleicht uns doch das Gefühl: das könnte integrierter, irgendwie (noch) vernetzter und intelligenter und damit effektiver und einfacher gehen!

Ein mögliches Szenario – und sicherlich das mittel- bis langfristige Ziel – könnte dabei aus dem Starten eines einzelnen Softwareprogramms (Softwareagent) bestehen, dem wir unseren „Plan“ mitteilen, und das diesen dann für uns recherchiert und als Ergebnis mit unter Umständen verschiedenen alternativen Lösungen aufwartet.

Ein weiteres, weniger komplex anmutendes und von „praktischer Eleganz“ gekennzeichnetes Szenario besteht in der Erkennung von Adressen, Telefonnummern und Terminen auf den Webseiten und deren Übernahme in ein persönliches Adressbuch oder einen Kalender.

Dieser Wunsch nach „integrierter“ und „einfacher“ setzt aber bestimmte Strukturen und Mechanismen voraus, die in Fachkreisen

unter dem Begriff „Semantisches Web“ (Semantic Web) in zunehmendem Maße das WWW von morgen prägen werden.

Dieses Buch will dem Leser diese Zukunft des WWW, seine Grundlagen und technischen Ausprägungen vorstellen und dabei zu ersten Gehversuchen mit der eigenen, privaten oder geschäftlichen Onlinepräsenz (z. B. Firmenwebseite oder dem eigenen Weblog) Lust machen und anleiten.

Dabei wird auf die Verwendung beliebter und gängiger Open-Source-Produkte auf der einen Seite und die Errungenschaften der „Semantic Web Community“ auf der anderen Seite besonderer Wert gelegt – sind beide doch treibende Kräfte (Innovatoren) und Ausprägung des Semantischen Webs.