

Semantic Web

Schon seit einigen Jahren fördert das W3C – insbesondere Tim Berners-Lee – die Idee des Semantic Web. Dabei stehen vor allem zwei Ziele im Vordergrund, die eng miteinander zusammenhängen: Zum einen ein verbesserter Austausch von Daten zwischen Anwendungen und zum anderen eine verbesserte maschinelle Auswertung von Daten.

Hintergrund

Das Internet ist öffentlicher Datenspeicher, Informationsquelle und Arbeitsplatz für Millionen von Menschen. Computer und Programme verwalten unzählige Datenbanken, Internetseiten und Kommunikationsplattformen. Diese Dienste basieren in aller Regel auf grundlegenden technischen Operationen, wie das Ausliefern von Dokumenten oder auch Fehlermeldungen sowie die Bereitstellung von Anwendungen für die Eingabe von Webinhalten. Programme können zwar auf die im Netz verfügbaren Daten zugreifen, aber es gibt nur begrenzte Möglichkeiten die verfügbaren Inhalte mit einem Mehrwert für die Benutzer auszuwerten.



Beispiel

Vorteile des Semantic Web lassen sich am besten anhand von Suchanfragen demonstrieren. Als ein typisches Beispiel werden hier die Schwierigkeiten einer Literaturwissenschaftlerin auf der Suche nach einem elektronisch vorhandenen Text skizziert:

Es gibt eine Übersetzung von Shakespeares Hamlet, die von Maik Hamburger stammt. Zwar wird die Fassung gerne an deutschen Schauspielhäusern gespielt, sie ist aber von keinem Verlag öffentlich publiziert. Die Literaturwissenschaftlerin versucht also, den Text im Internet zu finden und gibt bei einer Suchmaschine die Stichworte „Hamlet“ und „Hamburger“ ein. Nun ist es so, dass es in den USA eine Restaurantkette gibt, die „Hamburger Hamlet“ heißt (weil Hamlet im Englischen auch eine kleine Ortschaft bezeichnet). Überdies gibt es viele Aufführungen des Stücks Hamlet in der Stadt Hamburg, so dass Informationen über die Hamburgersche Übersetzung entweder gar nicht oder nur extrem schwer aufzufinden sind.

Gäbe es für die Wissenschaftlerin eine Möglichkeit, der Suchmaschine begrifflich zu machen, dass sie auf der Suche nach einem Buch mit dem Titel „Hamlet“ ist, das von einer Person namens „Hamburger“ verfasst bzw. übersetzt wurde, könnten die Restaurant-Kette und Berichte über Inszenierungen in Hamburg aus den Suchergebnissen ausgeschlossen werden.

Was wäre dafür nötig? Die Suchmaschine müsste „wissen“, dass ein Buch einen Titel und einen Autor hat und dass dieser Autor wiederum einen Namen trägt. Solche Informationen können Suchmaschinen derzeit aber nicht erfassen.

Ziele

Das Kürzel WYMIWYG steht für „What You Mean Is What You Get“ und lehnt sich an das aus der Textverarbeitung bekannte Kürzel WYSIWYG. Es steht für das zentrale Ergebnis aller Semantic-Web-Bemühungen. Der Nutzer soll bei Web-Recherchen das Ergebnis bekommen, das er tatsächlich gemeint hat, ohne durch zufällige technische Unzulänglichkeiten beeinflusst zu werden. Zudem ermöglicht die Modellierung von inhaltlichen Bezügen auch eine

automatische Erstellung von Verweisen (Hyperlinks).

Ansätze

Die Idee hinter dem Semantic Web, die Bedeutung von Daten maschinenlesbar aufzubereiten, um verbesserte Suchergebnisse, Hilfestellungen und Informationen bei der Arbeit mit Daten zu erhalten, geht bis in die 60er Jahre zurück und ist seither ein zentrales Feld im Bereich der Künstliche-Intelligenz-Forschung. Je nach Zielsetzung sind diese Ansätze in unterschiedliche Richtungen weiter entwickelt worden. Entsprechend stehen inzwischen unterschiedliche Werkzeuge zur Verfügung, die im Kontext des Semantic Web Bedeutung erlangen könnten.

Taxonomien

Taxonomien sind ein einfaches Werkzeug zur Strukturierung und inhaltlichen Auszeichnung von Daten. Sie werden seit langem im Bibliothekswesen und im Bereich von Lexika verwendet. Eine Taxonomie besteht aus einer Hierarchie von Schlagwörtern. Dabei bilden die Schlagwörter, die sich in der Hierarchie weiter oben befinden, die Oberbegriffe für die darunter eingeordneten Schlagwörter. Die eigentlichen Ressourcen werden immer mit dem präzisesten Schlagwort ausgezeichnet, weil die übergeordneten Schlagwörter aus der Hierarchie gefolgert werden können.

Taxonomien sind aus pragmatischen Gründe eine gute Wahl zur Verbesserung des Informationsangebotes, weil Benutzer mit Hierarchien bereits aus der Navigation mit Sitemaps vertraut sind und sie einen effektiven Weg bieten, Inhalte weitergehend zu strukturieren.

Ein großer Nachteil von Taxonomien ergibt sich dort, wo Ressourcen mehrere Elemente verbinden und nicht mehr eindeutig in einen Zweig des Kategoriensystems einzuordnen sind. Zudem ist es notwendig, die Taxonomie zu entwickeln und zu pflegen, was in einem dynamischen Wissensgebiet unter Umständen sehr aufwändig sein kann.

Wortnetze

Da es einige Wissensbereiche gibt, die sich nicht mit einfachen Taxonomien abbilden lassen, wurde bereits früh mit der Entwicklung von Netzen begonnen, die die Bedeutung von Schlagwörtern korrekt abbilden können. Sehr erfolgreich in einzelnen Anwendungen und entsprechend weit verbreitet sind Wortnetze nach dem Modell des Princeton Wordnet. Dabei werden synonyme Begriffe in so genannte Synsets zusammengefasst und mit semantischen Relationen untereinander verbunden.

So bilden beispielsweise die Begriffe „Auto“, „Automobil“, „Wagen“ und „PKW“ ein gemeinsames Synset. Ein weiteres Synset könnte aus „Fahrzeug“ und „Verkehrsmittel“ bestehen. Zwischen dem ersten und dem zweiten Synset besteht dann die Beziehung Unterbegriff-Oberbegriff (Hyponomie-Hyperonomie).

Wordnet bildet ein gutes Dutzend Relationen ab, die beschreiben, wie Begriffe und Konzepte zusammenhängen. Darunter fällt zum Beispiel die „Teil-Ganzes-Relation“, die es ermöglicht zu beschreiben, dass ein Kolben Teil eines Motors ist. Anhand dieser Relationen können komplexe Suchanfragen besser bearbeitet werden. Möchte jemand wissen wie ein Motor funktioniert, kann eine Liste aller Objekte generiert werden, die Teile eines Motors sind.

Weitere Informationen zu Wortnetzen bieten die Webseiten des Tübinger Projekts GermaNet.

Eine Zusammenstellung von Wortnetzen in zahlreichen Sprachen und die Lizenzen, unter denen sie stehen, bietet die von Samuel Chong (Pasadena City College) zusammengestellte "Complete Multilingual WordNet List by Language".

Objektorientierung

Objektorientierung wurde ursprünglich als Programmierparadigma entwickelt, für das unter anderem die Programmiersprache Java bekannt ist. Die Objektorientierung kennt dabei Klassen, die wiederum Eigenschaften und Methoden haben. So kann es z. B. eine Klasse „Auto“ geben, die die Eigenschaften „Hersteller“, „Höchstgeschwindigkeit“, „Sitzplätze“ und „Benzinverbrauch“ hat. Zusätzlich kann diese Klasse Methoden beinhalten, die beschreiben, was man mit einem Auto alles machen kann, wie bspw. Auftanken, Fahren, Reparieren.

Eine Klasse ist also eine abstrakte Beschreibung von Dingen. Von dieser abstrakten Beschreibung lassen sich konkrete Instanzen bilden, die die Eigenschaften und Methoden der Klasse übernehmen. Eine solche Instanz einer Klasse wird

Objekt genannt. So könnte es eine Instanz der Klasse Auto geben die »Aston Martin« heißt und dessen Eigenschaft „Sitzplätze“ den Wert „zwei“ hat.

Überdies kennt die Objektorientierung das Konzept der Vererbung. Dabei kann eine Klasse von einer anderen Oberklasse abgeleitet werden. Abgeleitete Klassen erben die Attribute und Methoden ihrer jeweiligen Oberklasse. Damit die abgeleitete Klasse keine reine Kopie ihrer Oberklasse ist, gibt es die Möglichkeit in den Unterklassen Attribute und Methoden zu ergänzen. So könnte die Klasse „LKW“ als zusätzliches Attribut „Ladefläche“ haben und als zusätzliche Methoden „Beladen“ und „Entladen“ vorsehen.

Beschreibungslogik

Logik-Formalismen wie die Aussagenlogik oder Prädikatenlogik erlauben das automatische Folgern von Aussagen. Dabei gilt es zunächst Regeln zu formulieren, wie z. B. „wenn A dann B, wenn B dann C“. Die Regeln werden um Zustandsbeschreibungen ergänzt wie etwa: „es gilt A“. Daraus ergibt sich dann automatisch C.

Dieses Beispiel stammt aus der Aussagenlogik. Umfassendere Formalismen wie die Prädikatenlogik steigern die Komplexität der formulierbaren Regeln und erlauben damit vielschichtigere Anfragen. Die Abbildung von logischen Zusammenhängen ermöglicht, auf unerwartete Fragen der Benutzer mit implizit gefolgerten Antworten zu reagieren.

Standards

Die bisher vorgestellten Konzepte zum Semantic Web haben gemeinsam, dass sie Wissen über inhaltliche Zusammenhänge in Form von Schlagwörtern repräsentieren, die durch bestimmte Formalismen verknüpft sind. Das Semantic Web tritt aber nicht zuletzt an, um eine weitere Ausbaustufe des Internets zu werden. Das Internet setzt sich allerdings nicht aus Schlagwortnetzen zusammen, sondern aus einer überwältigenden Menge von Webseiten, Texten, Bildern, Videos, Dokumenten - kurz Ressourcen, die jeweils über eine eindeutige Adresse, die URL auffindbar sind. Eine bessere Ordnung in dieses Ressourcenuniversum zu bringen ist eine zentrale Hoffnung, die mit Semantic-Web-Technologien verbunden ist.

Es stellt sich also die Frage, wie sich die geordneten Schlagwortnetze und die chaotischen Ressourcen des WWW zusammenbringen lassen. Ursprünglich sind zwei Standards angetreten, um die Ziele des Semantic Web zu verwirklichen. Auf der einen Seite steht das W3C mit seiner aufeinander aufbauenden Standard-Troika RDF, RDF-S und OWL. Auf der anderen Seite steht das Topicmap-Konsortium mit einem ISO-Standard, der ursprünglich aus dem Verlags- und Bibliothekswesen stammt und weit einfacher und intuitiver, dafür aber weniger aussagekräftig ist.

- Das Resource Description Framework (RDF) umfasst ein so genanntes Tripel aus Subjekt, Prädikat und Objekt, das es ermöglicht, neben Attribut-Wert-Paaren auch Beziehungen zwischen Ressourcen darzustellen.
- RDF-Schema (RDF-S) ist ein XML -Standard, mit dem für eine Klasse von ähnlichen Ressourcen wie z. B. Webseiten festgelegt werden kann, wie die Zuordnung von Eigenschaften erfolgen soll und welche grundlegenden Beziehungen zu anderen Eigenschaften bestehen. Ein Beispiel ist das Metadaten-Schema Dublin Core.
- Die Web Ontology Language (OWL) ist nach dem Prinzip der Objektorientierung aufgebaut und bietet sowohl die Möglichkeit, Attribute zu gestalten als auch Vererbung von Klassen abzubilden.

Anwendungen

Inzwischen gibt es eine ganze Reihe von Anwendungen zur Erstellung von Wissensbasen. Die meisten dieser Anwendungen erstellen Daten, die zu den Standards des W3C für semantische Auszeichnungen konform sind.

- Altova, ein Softwareunternehmen, das sich auf XML-Editoren spezialisiert hat, bietet seit einiger Zeit die SemanticWorks an. Als graphischer Editor soll es die professionelle Erstellung von W3C-konformen Ontologien ermöglichen.
- Eine Open-Source -Alternative zu SemanticWorks ist der Protege Editor, der in Stanford und ebenso wie SemanticWorks mit Hilfe von Java entwickelt wird. Der Protege soll ebenfalls die Erstellung von Ontologien durch qualifiziertes Personal ermöglichen.

Probleme und Ausblick

Obwohl das Semantic Web bereits seit vielen Jahren propagiert wird, ist bisher nur in einzelnen Vorzeige-Projekten

eine Realisierung zu erkennen. Die vorgestellten Semantic-Web-Standards sind zwar technisch gesehen Standards, aber noch weit davon entfernt, durch verbreiteten Gebrauch faktisch akzeptierte Standards darzustellen. In der Praxis ergeben sich erhebliche Probleme. Sie reichen von der Frage, wie detailliert man semantische Relationen modellieren soll, will oder kann, bis hin zur politischen Überlegung, ob es Autoritäten geben soll, die zentrale Bedeutungen festlegen.

Zunächst bedeutet die zusätzliche Anreicherung von Daten mit maschinenlesbaren Informationen auch zusätzliche Arbeit. Redakteure müssen neben der Erstellung des Artikels auch noch für die korrekte Verschlagwortung sorgen.

Des Weiteren spiegeln sich in der Auszeichnung mit Meta-Informationen auch Interessen wieder. So bot die Auszeichnungssprache HTML bereits in sehr frühen Versionen ein Element an, das Meta-Informationen über das jeweilige Dokument enthalten sollte. Es zeigte sich allerdings, dass Webseiten-Anbieter die Informationen im Meta-Tag verwendeten, um ihre Positionierung in den Suchmaschinen zu verbessern, unbeachtet, ob die Metainformationen inhaltlich richtig waren oder nicht. Daher gingen die großen Suchmaschinen dazu über, die Metainformationen zu ignorieren. Prinzipiell steht das Semantic Web vor demselben Problem – ohne eine Lösung dafür zu haben.

Die Konzepte des Semantic Web sind hochkomplexe Formalismen. Das bleiben sie selbst dann, wenn es gelänge, eine intuitive Anwendung zu entwickeln. In erster Linie bleibt der Aufbau semantischer Netze damit Experten vorbehalten. Stärker noch als Enzyklopädien definieren semantische Netze, was die Welt ist, was wahr ist und was falsch. Damit kommt einer sehr kleinen Gruppe von Menschen eine ausgesprochen große Definitionsmacht zu.

Eine alternative Herangehensweise stellen Social-Tagging-Systeme dar, die genutzt werden können, um aus einer großen Menge von individuell vergebenen Schlagwörtern dezentrale Ontologien (Folksonomies) aufzubauen. Daraus ergibt sich allerdings das umgekehrte Problem: Was einen Tag ausmacht wird über den Mittelwert der Masse aller, die Objekte mit diesem Tag versehen haben, definiert. Doch gibt es reichlich Beispiele für populäre Irrtümer.

Die Popularität von Tagging in diversen Web 2.0 -Anwendungen macht dennoch Hoffnung, dass zwei Grundannahmen des Semantic Web richtig sind: Zum einen geben sich viele Nutzer Mühe, Daten mit semantisch korrekten Metadaten anzureichern. Zum anderen scheinen sich Tags als sinnvolle und verständliche Suchmöglichkeit neben der Volltextsuche etabliert zu haben.

Letzte Änderung: 11.06.2015

Zitation

e-teaching.org (2015). Semantic Web. Zuletzt geändert am 11.06.2015. Leibniz-Institut für Wissensmedien:
https://www.e-teaching.org/technik/vernetzung/semanticweb/index_html. Zugriff am 20.10.2020

Barrierefreiheit [Direkt zum Inhalt](#) [Übersicht](#) [Erweiterte Suche](#) [Direkt zur Navigation](#) [Kontakt](#)