

Linked Open Data – Die Basis des Semantic Web

Die meisten von uns surfen jeden Tag im Netz. Wir posten Fotos vom Sonntagsbrunch, schauen Katzenvideos und googeln uns nach Lust und Laune durch verschiedene Websites. Aber wie funktioniert das World Wide Web eigentlich? Und wie kommt Google zu 69 200 000 Treffern, wenn ich „baby yoda“ suche?

Webseiten sind so aufgebaut, dass Menschen die Inhalte gut verstehen können. Ihre zugrundeliegenden Dokumente sind deshalb hübsch aufbereitet und oft mit Bildern ausgeschmückt. Das heißt aber nicht, dass die Daten auch von Computern formal verarbeitet werden können. Wenn zum Beispiel das Wort „Queen“ auf einer Webseite vorkommt, weiß der Computer nicht, ob es um die Musikgruppe oder die ehemalige Königin von England geht. Egal, wie viele Bilder dazu angezeigt werden. Deshalb kann auch bei so mancher Suchanfrage an Google ein schöner Blödsinn herauskommen. Obwohl nämlich ständig daran gearbeitet wird, die Suche zu verbessern, wird im Prinzip das Netz einfach nach den eingegebenen Stichworten durchsucht. Der Kontext, in dem sie vorkommen, kann dabei nicht berücksichtigt werden.

Die Idee des Semantic Web ist es, unser bestehendes World Wide Web zu ergänzen. Die Daten, die es schon gibt, sollen durch eindeutige Metadaten näher beschrieben werden. Die dahinterliegende Bedeutung wird also explizit gemacht. Wird auf einer Seite also der Werdegang der Band Queen beschrieben, baut man eine eindeutige Kennzeichnung ein, die auf genau diese Band verweist. Da-

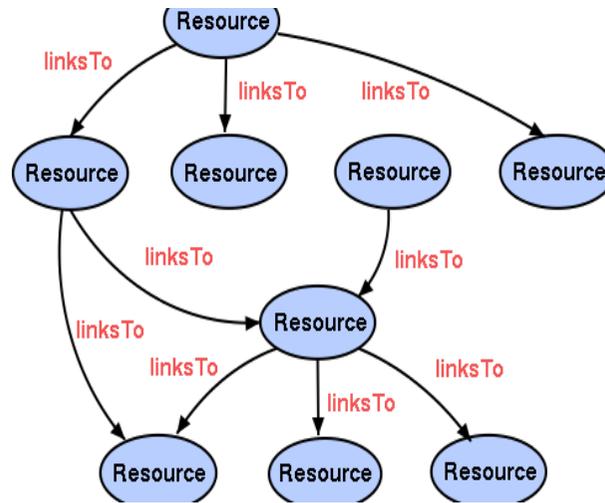


Abb.1: Verknüpfung von Inhalten im WWW (Bild: Eric Miller, W3C: <https://www.w3.org/2004/Talks/1116-semweb-em/slide9-0.html>)

durch weiß die Maschine, dass es hier um eine Band und nicht um die Königin von England handelt. Solche eindeutigen Kennzeichner heißen URIs (Uniform Resource Identifier) – auf Deutsch „einheitliche Bezeichner für Ressourcen“. Klingt nicht aufregend, ist aber in der Praxis sehr sinnvoll. Weil nämlich Computer mit diesen eindeutigen Kennzeichen viel mehr anfangen können als nur mit Wörtern. Die sind nämlich oft mehrdeutig, wie eben im Fall der Band und der Königin. Kommt derselbe URI auf zwei verschiedenen Seiten vor, weiß die Maschine aber, dass es hier um dasselbe Ding geht, und dich beide Seiten interessieren könnten, wenn du dieses Ding googelst.

Hier kommt Linked Open Data ins Spiel. Der Begriff vereint zwei wichtige Schlüsselwörter für den Bereich des Semantic Web. Einerseits sollen die Daten im World Wide Web nämlich „linked“, also miteinander verbunden, sein. Diese Verbindungen sind auch wieder – so wie die Daten selbst – maschinenlesbar. Dadurch kann der Computer den Kontext, der sich aus den verbundenen Daten ergibt, interpretieren. Und dadurch können wiederum Suchmaschinen verbessert werden.

Andererseits sollen die Daten offengelegt, also frei zugänglich sein. Das bedeutet, dass alle auf sie zugreifen und an den Verknüpfungen mitarbeiten können. Solange man sich dabei an die entwickelten Richtlinien und Vorgaben hält, kann ein Netz an verknüpften Daten entstehen, das riesiges Potenzial für Forschung, aber auch den privaten Nutzen birgt.

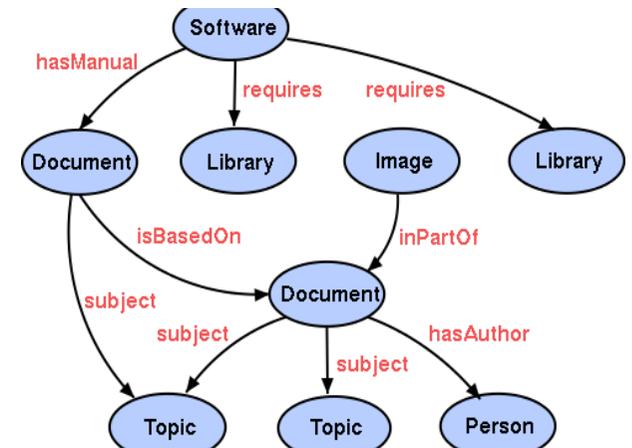


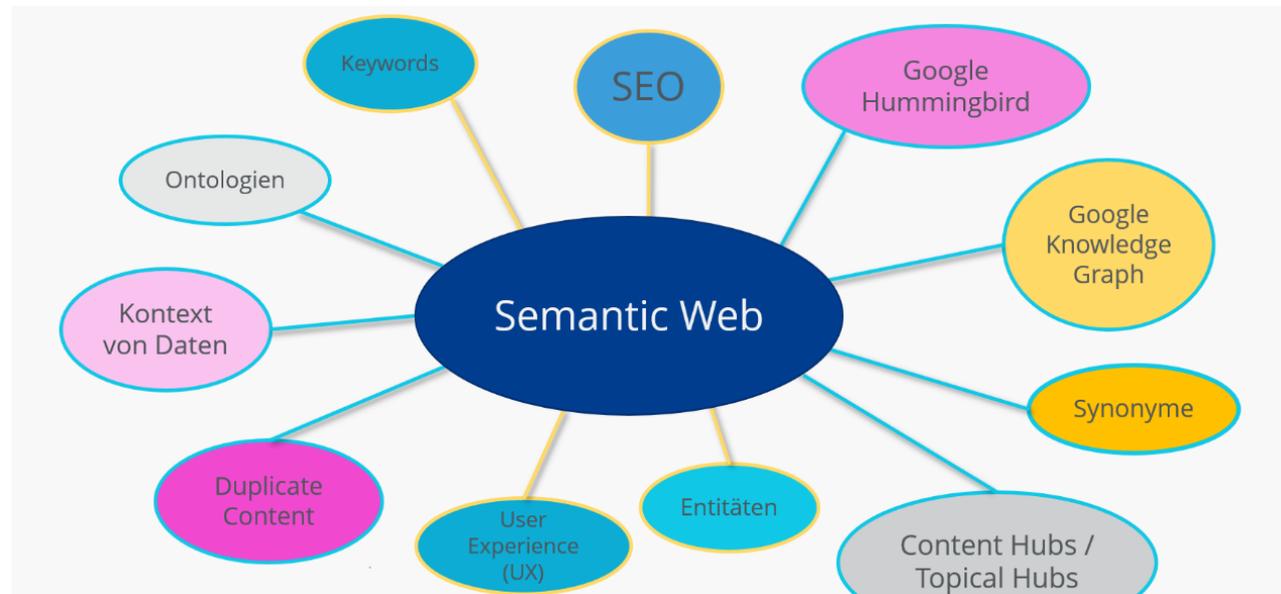
Abb. 2: Verknüpfung von Inhalten im Semantic Web (Bild: Eric Miller, W3C: <https://www.w3.org/2004/Talks/1116-semweb-em/slide10-0.html>)

Wissenschaftliche Literatur.

- Allemang, Dean; Hendler, James (2011): Semantic Web for the working ontologist. Effective modeling in RDFS and OWL. 2nd ed. Waltham: Morgan Kaufmann/Elsevier.
- Heath, Tom: <http://tomheath.com/blog/2009/03/linked-data-web-of-data-semantic-web-wtf/>
- Pollock, Jeffrey T. (2009): Semantic Web for dummies. Hoboken, NJ: Wiley.
- Rehbein, Malte (2017): Ontologien. In: Digital Humanities. Eine Einführung. Hrsg. von Fotis Jannidis, Hubertus Kohle und Malte Rehbein. Stuttgart: Metzler. S. 162-176.
- Sikos, Leslie (2015): Mastering structured data on the Semantic Web. From HTML5 Microdata to Linked Open Data. Berkeley: Apress.

Schlüsselbegriffe:

- SEO: Suchmaschinenoptimierung. Webinhalte werden so aufbereitet, dass Suchmaschinen sie leichter indexieren können.
- Google Hummingbird: Der Algorithmus, der in der Suchmaschine Google angewendet wird.
- Google Knowledge Graph: Eine Datenbank, aus der die Inhalte stammen, die in der Infobox neben den Suchergebnissen stehen.
- Synonyme: Zum Beispiel auf Basis von Bedeutung verwandte Wörter.



- Content / Topical Hubs: Zusammenhängende kuratierte Inhalte, die im Web angeboten werden.
- Entität: Im Semantic Web ist das die Kombination aus Identifikator („Ball“) und beschreibenden Informationen („Plastik“, „grün“).
- UX: 'User Experience' beschreibt wie ein Benutzer z.B. den Umgang mit einer Website erlebt.
- Duplicate Content: mehrfach vorhandene Inhalte.
- Ontologie: Sie beschreibt logische Zusammenhänge in einer geordneten, maschinenlesbaren Form.
- Keywords: Schlüsselbegriffe, mit denen der Inhalt von z.B. Webseiten beschrieben wird.

Abb. 4: Einige Schlüsselbegriffe des WWW / Semantic Web (Bild: <https://www.ionos.de/digitalguide/online-marketing/suchmaschinenmarketing/semantic-web/>)

