

Semantic-Web-Technologien

Zangerl, Lina Maria; linamaria.zangerl@plus.ac.at / Christopher Pollin; christopher.pollin@uni-graz.at

Das Semantic Web (Hitzler 2021), eine von Tim Berners-Lee geprägte Weiterentwicklung des World Wide Web, transformiert das Web miteinander verknüpfter Dokumente in ein Web miteinander verknüpfter Daten. In diesem erweiterten Kontext sollen die Technologien des *Semantic Web*, ihre Funktionen, Herausforderungen und möglichen Auswirkungen, insbesondere im Bereich der digitalen Editionen, betrachtet werden.

Die Idee einer Umgebung, in der Daten nicht nur zugänglich, sondern auch für Maschinen verständlich sind, steht im Mittelpunkt des *Semantic Web*. Diese Initiative zielt darauf ab, das Web in eine Plattform zu verwandeln, auf der Software-Agenten anspruchsvolle Aufgaben ausführen und auf der Grundlage der zugänglichen Daten autonome Entscheidungen treffen können (Berners-Lee, Hendler, and Lassila 2001). Diese Vision, die durch die Entwicklung der generativen KI noch verstärkt wird, zeigt, wie sich das *Semantic Web* und die generative KI synergetisch entwickeln bzw. dass die Ziele des *Semantic Web* unter Umständen überholt sind.

Die Architektur des *Semantic Web* beruht auf einer Reihe von Technologien, die oft als *Semantic Web Stack* zusammengefasst werden (Machado, Rocha Souza, and Simões 2019). Im Zentrum steht das *Resource Description Framework (RDF)*, ein Standardmodell für den Datenaustausch, das die Repräsentation von Informationen in Form von Subjekt-Prädikat-Objekt-Tripeln ermöglicht (Schreiber and Raimond 2014). Dies ermöglicht eine strukturierte, graphbasierte Darstellung von Daten. Das RDF-Schema (RDFS) ergänzt das RDF um die grundlegende Modellierung von Klassen und deren Beziehungen (Christophides 2009, 2425–28; Brickley and Guha 2014).

Eine weitere wichtige Komponente ist die *Web Ontology Language (OWL)*, die das Vokabular von RDFS erweitert und die Erstellung komplexerer Ontologien ermöglicht. Ontologien im *Semantic Web* sind formale, explizite Spezifikationen einer gemeinsamen Konzeptualisierung und zeichnen sich durch eine hohe semantische Ausdrucksfähigkeit aus (Feilmayr and Wöß 2016). Sie sind entscheidend für die Definition und Standardisierung von Konzepten und Beziehungen innerhalb spezifischer Domänen.

Eine weitere wichtige Technologie im *Semantic Web* ist die *Shapes Constraint Language (SHACL)* (Knublauch/Kontokostas 2017). Sie ermöglicht die Validierung von RDF-Daten gegen definierte Constraints und Strukturen, was für die Einhaltung und Überprüfung von Datenqualitätsstandards entscheidend ist. Dies ist insbesondere unter der Annahme der "*Open World Assumption*" relevant, die davon ausgeht, dass es niemals vollständiges Wissen über die im *Semantic*

Web enthaltenen Daten geben wird.

Kontrollierte Vokabulare und Thesauri spielen ebenfalls eine wichtige Rolle. Das *Simple Knowledge Organisation System (SKOS)* (Miles/Bechhofer 2009) ist ein wichtiges Beispiel für eine Ontologie, mit der strukturierte Vokabulare erstellt werden können. Solche Vokabulare sind für die Organisation und Klassifikation von Informationen, z. B. in Systematiken von Pflanzenarten, unerlässlich.

Für die Verwaltung von RDF-Daten werden spezielle Graphdatenbanken, sogenannte Triplestores, verwendet. Bekannte Beispiele sind Apache Jena, GraphDB, AllegroGraph und Blazegraph. Diese Datenbanken sind für die Speicherung, Abfrage und Manipulation von RDF-Daten optimiert. Mit SPARQL (Harris/Seaborne 2013), der *SPARQL Protocol And RDF Query Language*, steht eine umfangreiche Abfragesprache zur Verfügung, die es ermöglicht, komplexe Abfragen auf diesen Datenbanken durchzuführen und Graphen zu manipulieren und zu erzeugen.

Die Renaissance (Hitzler 2021) des *Semantic Web* wird durch *Knowledge Graphs* (Hogan and others 2021, 4) vorangetrieben, die komplexe Beziehungen zwischen Datenpunkten effektiv darstellen. Diese Graphen verknüpfen große Informationsmengen, ermöglichen tiefe, semantisch reichhaltige Analysen und erfüllen damit die ursprüngliche Vision des *Semantic Web*: ein vernetztes, semantisch durchdrungenes Informationssystem, das sowohl für Menschen als auch für maschinell lernende Systeme zugänglich ist.

Trotz seines großen Potenzials steht das *Semantic Web* auch vor Herausforderungen. Kritisiert (Swartz 2013) (Target 2018) werden seine Komplexität und die steile Lernkurve, die mit dem Verständnis und der Implementierung seiner Technologien verbunden ist. Hinzu kommen technische Herausforderungen bei der Verarbeitung großer Datenmengen und Infrastrukturprobleme, insbesondere bei der Verwendung von Open-Source-Software zur Verwaltung von RDF-Daten.

Das *Semantic Web* hat weitreichende Auswirkungen in verschiedenen Bereichen. In digitalen Editionen ermöglicht es beispielsweise die Darstellung semantischer Schichten in Dokumenten, wodurch diese als *Linked Open Data* zugänglich werden. Diese Fähigkeit erhöht die Reichhaltigkeit und Interaktivität digitaler Inhalte. Die Technologie hat auch potenzielle Anwendungen im Bereich des Wissensmanagements, der künstlichen Intelligenz und der Datenintegration und bietet neue Möglichkeiten für den Zugang, die Analyse und die Nutzung von Informationen.

Zusammenfassend lässt sich sagen, dass das *Semantic Web* einen transformativen Ansatz für die Strukturierung und Nutzung von Daten im Web darstellt. Obwohl es Herausforderungen in Bezug auf Komplexität und Akzeptanz gibt, bleibt sein Potenzial, die Interoperabilität von Daten und das maschinelle Verständnis von Webinhalten zu verbessern, groß. Mit der Weiterentwicklung der

Technologien und der Entwicklung benutzerfreundlicher Werkzeuge könnte das *Semantic Web* eine zentrale Rolle in der Evolution des Internets spielen, sofern dies nicht schon durch generative KI geschehen ist.

Das Semantic Web baut auf denselben Technologien wie das reguläre Internet auf, ergänzt diese aber. Basismodell ist das *Resource Description Framework* (RDF) (Cyganiak/Wood/Lanthaler 2014), das es ermöglicht, Ressourcen mittels einfacher *Triples* (Subjekt – Prädikat – Objekt) zu beschreiben. Auf RDF basieren auch die Sprachen *RDF Schema* (RDFS) (Brickley/Guha 2014) und die *Web Ontology Language* (OWL) (Motik/Patel-Schneider/Parsia 2012) die zur Beschreibung von Ontologien verwendet werden können. In einer Ontologie werden die Beziehungen zwischen Knoten und deren Typisierung beschrieben. Beispielsweise bietet *Friend of a Friend* (FOAF) (Brickley/Miller 2014) die Möglichkeit, Personen und deren Beziehungen zueinander zu beschreiben.

Ontologien gehen von einer *Open World Assumption* aus, was bedeutet, dass es grundsätzlich keine vollständige Kenntnis über die im *Semantic Web* enthaltenen Daten geben kann. Eine Validierung bestehender Datensätze gegen eine Ontologie ist also nicht möglich. Ob ein Datensatz konform zu einer Vorlage ist, kann daher mittels der *Shapes Constraint Language* (SHACL) (Knublauch/Kontokostas 2017) ermittelt werden.

Eine weitere technische Grundlage des *Semantic Web* sind kontrollierte Vokabulare bzw. Thesauri. Mit ihnen können strukturierte Klassifikationen, also etwa eine Systematik der Pflanzenarten, angelegt werden. Die maßgebliche Ontologie, nach der strukturierte Vokabulare angelegt werden können, ist das *Simple Knowledge Organisation System* (SKOS). (Miles/Bechhofer 2009) Ein weit verbreitetes Vokabular ist das ursprünglich von großen Suchmaschinen initiierte schema.org, das es ermöglicht, Webseiten semantische Informationen hinzuzufügen. In jüngerer Vergangenheit wurde der Anwendungsbereich auch auf Wissenschaft und Forschung ausgeweitet und für Metadaten des kulturellen Erbes im Rahmen von Europeana erprobt. (Freire/Charles/Isaac 2018).

Verwaltet werden RDF-Daten in *Triple Stores*, also Graph-Datenbanken. Einige verbreitete Produkte sind *Apache Jena*, *GraphDB* oder *AllegroGraph*. Mit der *SPARQL Protocol And RDF Query Language* (SPARQL) (Harris/Seaborne 2013) existiert eine umfangreiche Abfragesprache für diese Datenbanken, die auch die Bearbeitung und Erstellung von Graphen ermöglicht.

Literatur:

- RDF Schema 1.1 RDFS. URL: <https://www.w3.org/TR/rdf-schema/>
- FOAF Vocabulary Specification. URL: <http://xmlns.com/foaf/0.1/>

- RDF 1.1 Concepts and Abstract Syntax Concepts and Abstract Syntax. URL: <https://www.w3.org/TR/rdf11-concepts/>
- Freire, Nuno; Charles, Valentine; Isaac, Antoine: Evaluation of Schema.org for Aggregation of Cultural Heritage Metadata. In: The Semantic Web Lecture Notes in Computer Science. Cham: 2018, S. 225-239.
- SPARQL 1.1 Query Language SPARQL. URL: <http://www.w3.org/TR/sparql11-query/>
- Shapes Constraint Language (SHACL) SHACL. URL: <https://www.w3.org/TR/shacl/>
- SKOS Simple Knowledge Organization System Reference SKOS. URL: <http://www.w3.org/TR/skos-reference>
- OWL 2 Web Ontology Language Structural Specification and Functional-Style Syntax (Second Edition) OWL. URL: <https://www.w3.org/TR/owl2-syntax/>

Software:

Apache Jena, BlazeGraph, GraphDB, AllegroGraph

Verweise:

Semantic Web, Linked Open Data, Normdaten, RDF, SKOS

Themen:

Einführung, Annotation und Modellierung

Zitiervorschlag:

Zangerl, Lina Maria; Pollin, Christopher. 2021. Semantic-Web-Technologien. In: KONDE Weißbuch. Hrsg. v. Helmut W. Klug unter Mitarbeit von Selina Galka und Elisabeth Steiner im HRSM Projekt "Kompetenznetzwerk Digitale Edition". URL: <https://gams.uni-graz.at/o:konde.168>