

# Fünf Semantic Web Paradigmen

SEMINAR PAPER

25. Oktober 2014

---

**Simon Heimler**

heimlersimon@gmail.com

Master of Applied Research in Computer Science

**Informationssysteme**

Prof. Dr. Sabine Müllenbach

University of Applied Sciences Augsburg

---

## Abstract

---

Üblicherweise wird das Thema Semantic Web anhand des Semantic Web (Technologie) Stacks [1] erklärt. Dieser Artikel will einen alternativen Weg gehen: Nicht die technische Umsetzung steht im Vordergrund, sondern die Konzepte und Paradigmen hinter ihnen.

Dazu wurde eine subjektive Auswahl von fünf Semantic Web Paradigmen getroffen: (1) Mensch-Computer Kooperation, (2) die Mächtigkeit von Links (3) Graphenstruktur, (4) die Trennung von Fakt und Interpretation (5) und die Open World Assumption (Offene-Welt-Annahme).

Keines dieser Paradigmen ist ausschließlich im Semantic Web Kontext zu finden. Im Gegenteil: Alle diese Konzepte haben ihren Ursprung in anderen Disziplinen. Selbst wenn der Leser also nicht die ganze Semantic Web Vision „kauft“, können diese Paradigmen dennoch eine wertvolle Bereicherung sein, da sie interessante Lösungen zu aktuellen Problemen aufzeigen.

Wenn man die Paradigmen zusammenführt bilden sie (zusammen mit weiteren) die Grundlagen der Semantic Web Vision.

**Keywords:** Semantic Web, Linked Data, Mensch-Computer-Kooperation, Open World Assumption, Graphenstruktur

## Abbreviations

---

SW	Semantic Web
MCK	Mensch-Computer-Kooperation
OWA	Open World Assumption

# 1 Einleitung

---

Ein bekanntes Zitat, dass nachträglich diversen Autoren zugeschrieben wird, besagt: *„Wer als Werkzeug nur einen Hammer hat, sieht in jedem Problem einen Nagel.“* Der effizienten Lösung von Problemen tut dies natürlich nicht gut. Wie kann man diesem Problem also entgehen?

Die Lösung liegt nahe: Man beschäftigt sich mit anderen, neuen Werkzeugen. Das Wort Werkzeug ist hier natürlich metaphorisch gesprochen: Oberflächlich gesehen könnte damit das einem zu Verfügung stehende Handlungsrepertoire gemeint sein. Doch hinter diesem stehen unsere (oft festgefahrene) Paradigmen, also Denkweisen und Weltanschauungen (die unser Handeln ursächlich bestimmen).

Diese Argumentation sollte auch darlegen warum sich dieser Artikel mehr auf das „Warum?“ konzentriert und weniger auf das „Wie?“. Die konkreten technischen Werkzeuge ändern sich. Doch die grundsätzlichen Ideen hinter ihnen sind beständiger und für das Verständnis wichtiger.

Das Ziel dieses Artikels ist es also dem Leser einige Paradigmen aus der Semantic Web Community vorzustellen. Sie zeigen alternative Herangehensweisen zu aktuellen Problemen auf, die bisher relativ unbekannt sind.

## 2 Vorstellung der Paradigmen

---

### 2.1 Mensch-Computer-Kooperation

---

Es ist heute nicht mehr zu verleugnen, dass es einige Gebiete gibt in denen Computer erheblich besser und effizienter arbeiten als Menschen. Aber man kann auch umgekehrt argumentieren: Trotz allen Fortschritten in der KI Forschung gibt es viele Bereiche in denen Menschen nicht ersetzt werden können – oder sollten.

Das Konzept der Mensch-Computer-Kooperation bietet eine weitere Sichtweise: Menschen haben bestimmte Stärken und Schwächen - und Computer die ihre. Wenn beide Seiten miteinander auf produktive Weise zusammenarbeiten und dies berücksichtigt wird, kann das Ergebnis beide Seiten im Alleingang weit übertreffen. [2]

Das Semantic Web kann als Mensch-Computer-Kooperation Initiative für das Web verstanden werden. Wenigen wird dies bewusst sein, doch das aktuelle Web ist für Menschen optimiert und für Maschinen oft nur sehr schwierig und missverständlich zu interpretieren.

Um eine produktivere Kooperation zu ermöglichen müssen Webseiten also so verfasst und erstellt werden, dass sie sowohl von Menschen als auch Maschinen gut verstanden werden können. Metadaten lösen dieses Problem nicht, da sie nur die Dokumente und Dateien an sich beschreiben, nicht aber dessen Inhalt.

Das Semantic Web will dieses Problem durch semantische Annotation lösen. Semantisch bedeutet in diesem Kontext, dass die tatsächliche *Bedeutung* und der Inhalt auf maschinenlesbare Weise erfasst werden. Dies ist natürlich ein sehr ambitioniertes Vorhaben und es bleibt noch abzuwarten in welchem Umfang und Qualität es sich erfüllen wird.

Warum sollte man sich also diesen Aufwand machen?

### 2.1.1 Aktueller Stand

Semantische Annotation ist der Teil des Semantic Webs der bis jetzt die beste Annahme gefunden hat. Viele große Webunternehmen haben die letzten Jahre begonnen semantische Annotationen zu fördern und zu verarbeiten. Damit wird die Annotation für Webentwickler interessant bis wichtig, da sie bessere SEO und Integration mit diesen Anbietern verspricht. Die Nutzer profitieren dann wiederum von den neuen Angeboten und Services die aufgrund dieser Technologie möglich oder besser geworden sind.

Schema.org [3] ist Projekt von Google, Yahoo, Microsoft, etc., dass ein gemeinsames Vokabular für die Beschreibung von Dingen und Vorgängen im Internet definiert. Facebook hat einen eigenen Standard entwickelt, den OpenGraph [4].

Auch Forschung und Industrie haben die Semantic Web Technologien aufgenommen um übergreifende Standards zu schaffen, wie etwa das SKOS Modell [5] zur Verwaltung von Wissensbeständen.

### 2.1.2 Technischer Hintergrund

Aktuell gibt es viele unterschiedliche Semantic Web Datenserialisationsformate, einige von ihnen vom W3C standardisiert. Größere Verbreitung haben RDFa [6] und Microdata [7], die beide auf XML basieren. Steigende Verbreitung hat aktuell JSON-LD [8], das auf dem einfachen JSON Datenformat aufbaut. Als textbasiertes Format ist Turtle [9] zu erwähnen.

Alle diese Formate teilen sich ein zugrundeliegendes Daten-Konzept: RDF [10]. Auf dieses wird im Abschnitt 2.2 tiefer eingegangen.

## 2.2 Die Mächtigkeit von Links

Hyperlinks sind nicht erst mit dem Web erfunden worden. Sie haben eine lange Geschichte die mindestens bis zur Memex zurückreicht und waren wichtiger Bestandteil der in den 60er Jahren entstehenden Hypertext Systeme [11].

Der Erfinder des WWW, Tim-Berners Lee entschied sich dafür eine vereinfachte Version der Links zu verwenden die nur unidirektional sind und bei denen die Relation zwischen Quelle und Ziel kaum eine Rolle spielt. Die unidirektionale Natur der Verknüpfungen ermöglichte die dezentrale Struktur des Internets, da keine zentrale Datenbank aller Verknüpfungen nötig war und Links ohne vorherigen Konsens gesetzt werden können.

Rückblickend wird diese Entscheidung für den Erfolg des Webs verantwortlich gemacht. [12]

Wenn wir von Links sprechen, meinen wir meist nur das Ziel dieser Verknüpfung, die URL (Unique Resource Identifier): Eine eindeutige Adresse die eine Ressource identifiziert.

Wir haben also eine vereinfachte und damit nicht so mächtige Version von Links im Web. Das Semantic Web setzt an dieser Stelle ein und baut das Konzept der Links und URLs weiter aus:

Links können als Tripels verstanden werden: Quelle, Relation und Ziel. Im Semantic Web wird daraus das Konzept von RDF, dass diese in eine minimale grammatische Aussage umformuliert: Subjekt, Prädikat und Objekt. Die Relation bekommt damit eine zentrale Rolle: Erst durch sie bekommt die Aussage auch eine Bedeutung.

URLs werden im Semantic Web allerdings nicht nur verwendet um Adressen zu Webseiten anzugeben. Sie können auch auf abstrakte Dinge oder Beziehungen verweisen. Der Begriff „Resource“ wird hier also ausgeweitet. Dadurch ist es möglich auch über Dinge, Konzepte und Beziehungen zu reden und dennoch eine eindeutige Referenz zu haben, die auch über Grenzen, wie verschiedene Webseiten oder Datenbanken, hinweg funktionieren kann.

## 2.3 Graphen- und Netzwerkstrukturen

---

- 
- Triple links result in a graph structure [13]
- Graphs – one structure to rule them all. Merging of Structures [14]
- Can grow dynamically / organically
- The nature of things is often a graph, trees and lists are mostly simplifications / abstractions.

### 2.3.1 Technischer Hintergrund

---

- Triplestores, Graph databases
- SPARQL, Gremlin, Cypher

## 2.4 Trennung von Fakt und Interpretation

---

- A graph can be a collection of facts, there is no explicit need for a schema.
- However something like a schema can be put on top. Something even more powerful: An ontology
- Background in AI research and mathematical logic
- Support for inference and reasoning

- Ontology and Facts are strictly separated. It is possible to have one fact dataset and multiple ontologies, with different interpretations.
- Ontology / Schema can grow with the data.
- Relational: Schema first, without a schema it is not possible to organize/store the data. With graphs this is not mandatory.

### 2.4.1 Technischer Hintergrund

---

- RDFS, OWL, (Rules??)

## 2.5 Die Open World Assumption

---

- Maybe the most complicated concept
- Deals with incomplete knowledge
- Tries to avoid contradictions, infers new knowledge instead
- Interesting concept for aggregated knowledge bases with data from various sources
- (Fuzziness, Trust, Provenance)

## 3 Schlussbetrachtung

---

- Semantic Web Technologies have interesting solutions for current problems which are mostly unsolved.
- Standards and Implementations will change but the problems behind not.
- Understanding the concepts behind may prove to be useful, no matter if the Semantic Web Vision gets the adoption it needs.

## References

---

- [1] Tim Berners-Lee, *The next web*. Available: [http://www.ted.com/talks/tim\\_berniers\\_lee\\_on\\_the\\_next\\_web.html](http://www.ted.com/talks/tim_berniers_lee_on_the_next_web.html) (2013, Oct. 08).
- [2] S. Sankar, *The rise of human-computer cooperation*. Available: [http://www.ted.com/talks/shyam\\_sankar\\_the\\_rise\\_of\\_human\\_computer\\_cooperation?language=en](http://www.ted.com/talks/shyam_sankar_the_rise_of_human_computer_cooperation?language=en) (2014, Oct. 31).
- [3] Google Inc, Yahoo Inc, Microsoft Corporation and Yandex, *schema.org*. Available: <http://schema.org/> (2014, Oct. 24).
- [4] Facebook, *Open Graph protocol*. Available: <http://ogp.me/> (2014, Oct. 25).
- [5] Antoine Isaac, Ed Summers, *SKOS Simple Knowledge Organization System Primer*. Available: <http://www.w3.org/TR/2009/NOTE-skos-primer-20090818/> (2014, Oct. 31).
- [6] Manu Sporny, *HTML+RDFa 1.1*. Available: <http://www.w3.org/TR/2013/REC-html-rdfa-20130822/> (2013, Oct. 14).
- [7] Ian Hickson, *HTML Microdata*. Available: <http://www.w3.org/TR/microdata/> (2014, Oct. 24).

- [8] Manu Sporny et al, *JSON-LD 1.0*. Available: <http://www.w3.org/TR/json-ld/> (2014, Oct. 24).
- [9] Gavin Carothers and Eric Prud'hommeaux, *Turtle*. Available: <http://www.w3.org/TR/2013/CR-turtle-20130219/> (2013, Oct. 18).
- [10] Eric Miller and Frank Manola, *RDF Primer*. Available: <http://www.w3.org/TR/2004/REC-rdf-primer-20040210/> (2013, Oct. 14).
- [11] T. H. Nelson, "Complex information processing," in *the 1965 20th national conference*, pp. 84–100.
- [12] Hendrik Arndt, *Integrierte Informationsarchitektur: Die erfolgreiche Konzeption professioneller Websites*. Springer, 2006.
- [13] s. V. Wikipedia-Autoren, *Resource Description Framework - Wikipedia, the free encyclopedia*. Available: <http://en.wikipedia.org/w/index.php?oldid=580735405> (2013, Nov. 12).
- [14] P. Hitzler, *The semantic web: Proceedings*. Berlin, Heidelberg, New York, NY: Springer, 2007.