Five Semantic Web Paradigms

SEMINAR PAPER

24th October 2014

Simon Heimler
heimlersimon@gmail.com
Master of Applied Research in Computer Science

Informationssysteme
Prof. Dr. Sabine Müllenbach
University of Applied Sciences Augsburg

Abstract

Semantic Web Technologies and even the Linked Data initiative had a rather slow adoption rate over the past years. This may be caused by the introduction of many unfamiliar and sometimes complicated concepts. This article takes does not explain concrete technology or implementation details to achieve a Semantic Web. It takes a step backward and gives a brief overview over a selection of five fundamental Semantic Web paradigms: (1) Human-Computer Cooperation, (2) Identification through URIs, (3) Graph and Network Structure, (4) the separation of fact and interpretation and (5) the Open World Assumption (OWA).

These paradigms are not bound to be used only in the context of Semantic Web Technologies. In fact, each of the paradigms also exists independent from Semantic Web Technology. So even if one does not buy into the full Semantic Web Stack and Vision, those paradigms can still be a useful addition to one's tool- and mindset.

If put all together, these paradigms (together with some more) build the conceptual foundation of Semantic Web Technologies. Understanding them is crucial to see why the Semantic Web Vision came up in the first place.

Keywords: Semantic Web, Linked Data, Human-Computer Cooperation, Open World Assumption, Graphs, Networks.

Abbreviations

SW	Semantic Web
HCC	Human-Computer cooperation
OWA	Open World Assumption

1 Introduction

Usually the Semantic Web is explained through the Semantic Web Stack [1], beginning at the bottom with the most fundamental technologies up to the more sophisticated ones on the top. The concepts are explained along the currently used standards and their implementations. This article tries a different approach: A (subjective) selection of five semantic web core concepts will be explained on a more abstract and isolated level - as paradigms. The focus will be on the why and not so much the how, since this changes over time.

As the reader will hopefully notice, all of those paradigms are interesting and useful concepts by their own right. They are note exclusive to the Semantic Web Community, in fact many of those concepts existed before the Semantic Web.

The author hopes that reading about those paradigms broadens the reader's mind- and toolset. Maybe some of them can be employed in completely different contexts. Some problems sometimes require unconventional solutions and the Semantic Web may offer some interesting ideas here.

2 A Selection of Five Semantic Web Paradigms

2.1 Human-Computer Cooperation

Undeniable there are things that computers are much better than humans. But it goes the same way round: Despite much progress in Artificial Intelligence research there are a lot areas where Humans can't (or shouldn't) be replaced, especially if it comes to intuition or responsibility.

This line of thought leads us to the concept of Human-Computer cooperation. Its main philosophy is: Computers have specific strengths and so do humans. If they both can work together in a productive manner the result can excel the performance of either.

The Semantic Web can be thought as a Human-Computer cooperation initiative for the Web. Right now the Web is optimized toward humans. For some this might be surprising, but machines have a hard time interpreting the actual *meaning* of web documents.

To provide a basis for successful Human-Computer cooperation, web documents have to be written in a manner so that it can be well understand by both humans and machines. This is done through semantic annotation, which adds an additional layer of machine-optimized data.

Now that the computer can interpret the information on a website, it can assist humans in various ways, that would be tedious or error prone for them: Aggregation (and recombination) of information and knowledge. A more seamless integration of different websites and their offerings. The possibilities are manifold.

2.1.1 Current State

Semantic annotation is currently the Semantic Web feature that has the best adoption rate. Many big web companies added support for them. There is a shared effort project between Google, Yahoo, Microsoft, etc. to define a web-wide vocabulary to describe facts, called schema.org [2]. Facebook decided to create its own standard, called OpenGraph [3].

2.1.2 Technical Details

Currently there are many different and competing Semantic Web data serialization formats, some of them W3C standardized. The most important of them are RDFa [4] and Microdata [5] which are both XML based. There is a JSON based format called JSON-LD [6] and the text based Turtle [7]. They all share a common concept that is called RDF [8] The actual data is stored as triples of subject, predicate and object. Every element of the triple is an URL.

2.2 Identification through URIs

URI's are one of the main components of the web. The Semantic Web takes the concept of links a step further: Not just websites have URL's but also things, like persons or relationships.

If a thing has a unique identifier through an URL, different websites or datasets can talk about the same thing across borders.

- Links always had this potential
- There is indication that Tim Berners-Lee had envisioned the web that way from start
- Computers can follow the Link and get more information there. This makes the web traversable in an information/data oriented way

This concept is especially powerful if combined with the concepts of graph structures.

2.3 Graph and Network Structures

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

2.3.1 Technical Details

- Triplestores, Graph databases
- SPARQL, Gremlin, Cypher

2.4 Separation of fact and 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 Technical Details

- RDFS, OWL, (Rules??)

2.5 The 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 Conclusion

- 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, *Semantic Web, and Other Technologies to Watch.* Available: http://www.w3.org/2007/Talks/0130-sb-W3CTechSemWeb/#(24) (2014, Oct. 24).

[2] Google Inc, Yahoo Inc, Microsoft Corporation and Yandex, *schema.org.* Available: http://schema.org/ (2014, Oct. 24).

[3] Facebook, *Open Graph protocol.* Available: http://ogp.me/ (2014, Oct. 25).

[4] Manu Sporny, *HTML+RDFa 1.1.* Available: http://www.w3.org/TR/2013/REC-html-rdfa-20130822/ (2013, Oct. 14).

[5] Ian Hickson, *HTML Microdata.* Available: http://www.w3.org/TR/microdata/ (2014, Oct. 24).

[6] Manu Sporny et al, *JSON-LD 1.0.* Available: http://www.w3.org/TR/json-ld/ (2014, Oct. 24).

[7] Gavin Carothers and Eric Prud'hommeaux, *Turtle.* Available: http://www.w3.org/TR/2013/CR-turtle-20130219/ (2013, Oct. 18).

[8] Eric Miller and Frank Manola, *RDF Primer.* Available: http://www.w3.org/TR/2004/REC-rdf-primer-20040210/ (2013, Oct. 14).