Semantic Web

Author:
Felix Kostenzer

Supervisor:
Prof. Dr. Georg Moser

Contents

1 Introduction **1**
1.1 Web 2.0 1
1.2 Web 3.0 1

2 Resources **2**
2.1 Identification 2
2.2 Triple 2

3 Ontology **3**
3.1 Definition 3
3.2 OWL 3

4 Conclusion **4**

References **5**

1 Introduction

In the following section, the reader will be introduced into the conceptual difference between the WWW, as we know it today, and the Semantic Web.

1.1 Web 2.0

The World Wide Web as we know it today offers lots of information about nearly all aspects. For instance take a look at the well known online encyclopedia Wikipedia. It is designed to provide tons of information written by humans for humans. This means that information is presented in a way, may it be text an image or whatever, so that it can be understood by humans. By taking a closer look it will become clear that this is more than unfavorable if an automated system e.g. a computer tries to comprehend what's going on there. Moreover an algorithm or system can not reliable determine the meaning of such chunk of data. This is a major drawback of the current WWW as automated processing of it's content is very limited.

1.2 Web 3.0

The idea behind the Semantic Web is to add an additional layer to the existing web which provides semantic information about it's content. By utilizing this extra data an automated system can combine 'things' by their meaning and generate 'new knowledge' based on existing one. At this point do not confuse 'meaning' with the presentation of web-content which has only expressive power to human beings.

"*The Semantic Web will enable machines to COMPREHEND semantic documents and data, not human speech and writings.*"(Berners-Lee et al., 2001)

In order to provide an example which exposes some of the Semantic Web's power take a look at the following: "*Suppose you wish to find the Ms. Cook you met at a trade conference last year. You don't remember her first name, but you remember that she worked for one of your clients and that her son was a student at your alma mater. An intelligent search program can sift through all the pages of people whose name is "Cook" (sidestepping all the pages relating to cooks, cooking, the Cook Islands and so forth), find the ones that mention working for a company that's on your list of clients and follow links to Web pages of their children to track down if any are in school at the right place.*"(Berners-Lee et al., 2001)

2 Resources

This section introduces the fundamental stones upon which the Semantic Web is built on.

2.1 Identification

Recall the usage of links within the context of HTML. These sort of references are very useful as they allow to make an uni-directional connection from A to B. For instance http://www.example.com/index.html includes an image hosted on another server so the obvious way is to do external linkage. Within the Semantic Web we need a bit more than that. The reason for this is that it must be possible to distinguish between legacy web-content and so-called objects. Those objects are the semantic representation of real-life things like a person, a car or anything else. However there are two commonly used solutions for this issue. First one are 303-URIs and second one are Hash-URIs[1]. By using this way of identifying 'things' the Semantic Web has the ability to effectively separate representation of data and their semantic description.

2.2 Triple

A so called 'triple' represents a fundamental statement. It consists of subject, predicate and an object. The subject and object can be seen as a reference to real-life objects and the predicate defines the relationship between those two. For instance <Adam> <hasSister> <Eva> would be such a triple. After this point, the relationship between Adam and Eva is clearly defined such that an automated system can understand it. In order to properly describe such sentences in a semantic manner the Semantic Web draws on the so called RDF & RDF Schema[2]. Via this framework it is possible to describe such relations as well as objects in an abstract manner. As like in object oriented programming languages the RDF Schema allows to define heritage, properties, value ranges and so on. Recall the example from above. It would make sense to define a class named <Person> which holds a set of generic attributes that are acceptable for all 'Persons'. The concrete objects <Adam> and <Eva> will then simply be instances of the <Person> class with the correct values set.

[1] http://www.w3.org/TR/2008/NOTE-cooluris-20081203/
[2] http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/

3 Ontology

In the following section the term 'ontology' is introduced as well as OWL.

3.1 Definition

Before hopping into ontologies let's reconsider the fundamentals which are essential for any further step.

1. URIs which are used to uniquely identify a web-resource or semantic data aka real-life object.
2. "*RDF is a datamodel for objects ("resources") and relations between them, provides a simple semantics for this datamodel, and these datamodels can be represented in an XML syntax.*"(McGuinness and van Harmelen, 2004)
3. "*RDF Schema is a vocabulary for describing properties and classes of RDF resources, with a semantics for generalization-hierarchies of such properties and classes.*"(McGuinness and van Harmelen, 2004)

Having all those three things set up is important but not enough in the context of the Semantic Web. An important factor that is missing in the equation is the so called ontology.

"*In philosophy, an ontology is a theory about the nature of existence, of what types of things exist; ontology as a disciplinestudies such theories. Artificial-intelligence and Web researchers have co-opted the term for their own jargon, and for them an ontology is a document or file that formally defines the relations among terms. The most typical kind of ontology for the Web has a taxonomy and a set of inference rules. The taxonomy defines classes of objects and relations among them.*"(Berners-Lee et al., 2001)

Speaking in simple terms, an ontology can be seen as an extension to what RDF and RDF Schema provide but with additional expressive power. By utilizing this, web-content can be described based on logic and therefore allows automated systems to interpret such content.

3.2 OWL

OWL aka Web Ontology Language is used to describe relationships among objects. Some of the important aspects of what can be done via OWL[3] are:

[3] http://www.w3.org/TR/2004/REC-owl-guide-20040210/

1. As like in XML, OWL allows you to declare namespaces. These are particularly useful if you want to introduce a certain set of vocabulary within the document. Furthermore it is possible to include all assertions from other ontologies. This increases the re-useability as defining the same statements over and over again simply does not make sense.

2. As like in RDF, the language allows to define classes including heritage, properties, literals and so on. But there is a major difference. The properties can and should be assigned with additional meaning like transitivity, symmetry etc.

3. Different ontologies might share similar or equal classes, individuals and even properties. In order to enhance re-usability there are certain keywords which allow to indicate such equivalence or anti-valence.

4. Restrictions for class members based on set operations like intersection, union or complement.

4 Conclusion

The Semantic Web is indubitable a step forward for the WWW as such. Connecting data based on their meaning will not only enhance search engines but also tighten the interlink between knowledge. While offering new possibilities and expressive power there are some drawbacks though. I personally think that, at least due to a certain extend, the current WWW became so successful because it is easy to contribute something for nearly everyone. Don't get me wrong here, I do not mean posting something on Twitter or Facebook but to e.g. host a website or to write an article on Wikipedia. I suppose that the Semantic Web is just too complex for the 08/15 user. You cannot really demand from an average user, who is only used to e.g. Adobe Dreamweaver or Internet Explorer, to write correct ontologies including all 'fancy' things like set operations, inference rules etc. This won't simply do, and therefore the question is who will write it then? As it cannot be generated by an automated system it must be done by a human which is expensive and of course time intensive. Furthermore how is it possible to verify the trustworthy of the supplied semantic data? There are still lots of unanswered questions that need to be answered first. Without those issues solved, the Semantic Web will most likely remain an 'idea' instead or replacing the current WWW.

References

Tim Berners-Lee, James Hendler, Ora Lassila, et al. The semantic web. *Scientific american*, 284(5):28–37, 2001.

Jeremy J. Carroll and Graham Klyne. Resource description framework (RDF): Concepts and abstract syntax. W3C recommendation, W3C, February 2004. http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/.

Richard Cyganiak and Leo Sauermann. Cool URIs for the semantic web. W3C note, W3C, December 2008. http://www.w3.org/TR/2008/NOTE-cooluris-20081203/.

AnHai Doan, Jayant Madhavan, Pedro Domingos, and Alon Halevy. Learning to map between ontologies on the semantic web. In *Proceedings of the 11th international conference on World Wide Web*, pages 662–673. ACM, 2002.

Patrick Hayes. RDF semantics. W3C recommendation, W3C, February 2004. http://www.w3.org/TR/2004/REC-rdf-mt-20040210/.

Deborah L. McGuinness and Frank van Harmelen. OWL web ontology language overview. W3C recommendation, W3C, February 2004. http://www.w3.org/TR/2004/REC-owl-features-20040210/.

Eric Miller and Frank Manola. RDF primer. W3C recommendation, W3C, February 2004. http://www.w3.org/TR/2004/REC-rdf-primer-20040210/.

Nigel Shadbolt, Tim Berners-Lee, and Wendy Hall. The semantic web revisited. *IEEE Intelligent Systems*, 21(3):96–101, May 2006. ISSN 1541-1672. doi: 10.1109/MIS.2006.62. URL `http://dx.doi.org/10.1109/MIS.2006.62`.

Chris Welty, Deborah L. McGuinness, and Michael K. Smith. OWL web ontology language guide. W3C recommendation, W3C, February 2004. http://www.w3.org/TR/2004/REC-owl-guide-20040210/.