Semantic Web

Gabriel de L. Rabelo[1]

[1]*Department of Computer and Information Sciences, Temple University, Philadelphia, PA.*

***Abstract**: This paper discusses the Semantics Web from the Artificial Intelligence perspective. The Semantic Web is an extension to the existing Web providing a standardization to data representation giving it meaning and making it understandable to computers, which enables them to perform several sophisticated tasks.*

1. Introduction

The way the web currently works is, in a concise manner, as a web of documents. Those documents are given addresses to make them accessible and contain links to other documents. The degree of structured data on this approach is fairly low.

On a common web search engine, a query as "*How many movies has Morgan Freeman been in?*" would result in a list of pages that contains the words on the query, demanding the user to look through several links in order to get the answer, instead of getting it in the first place. This happens because the search was performed on the documents instead of the data itself, therefore, as an answer, the user gets a list of documents as an answer.

The current web is evolutioning in a way that is overcoming its initial design and thus provoking some issues, e.g. the search engines output. The amount of information currently available on the web is enormous and it is becoming more difficult to find relevant information to answer a query, which can be proven by all the search engines outputs that are not accurate to the query. Search engines encounter difficulties relying on matching the keywords on the query to words that appear on web documents. Natural language ambiguity, synonyms and other language aspects aggravate this problem. And this is just one of many issues that are rising.

The tendency now is building a web that supports new requirements as locating information by changing the searches from simple keyword matching to more sophisticated semantic techniques, automating tasks by making the machine to filter and process the data before the user's consumption along with more complex tasks, and web services [13].

As described by Tim Berners-Lee [1], **"**The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation.". Representing the third generation of the web, the Web Semantics is a Web of Data. It is concerned with the meaning of the data, and this is its fundamental difference in relation to the current web. It focus on building a web that can be understood by computers enabling them to perform sophisticated tasks on the users behalf.

The Semantic Web allows the search to combine data from many different sources and use reasoning over it in order to answer the query. Taking the same example mentioned above, if one wanted to know about an specific actor, the search engine could formulate the answer using information on the Linked Movie Data Base (http://www.linkedmdb.org/), a Semantic Web database dedicated to movie-related information, about the movies this actor is associted with. In addition, it could associate the actor's information with DBpedia, the Wikipedia formated for Semantic Web, to be able to include his biography, career history and facts. This way, the answer to the query would be the actual information about the actor instead of links to pages that are related to him.

2. The Semantic Web

According to the W3C, which holds the standardization for the Semantic Web in the context of Web 3.0, the Semantic Web is about two things: Common formats for integration and combination of data and language for recording how data relates to real world objects [3].

Created by Tim Berners-Lee, the Semantic Web Stack is used as a representation of the architecture and the hierarchy of languages and technologies in use on the Semantic Web.

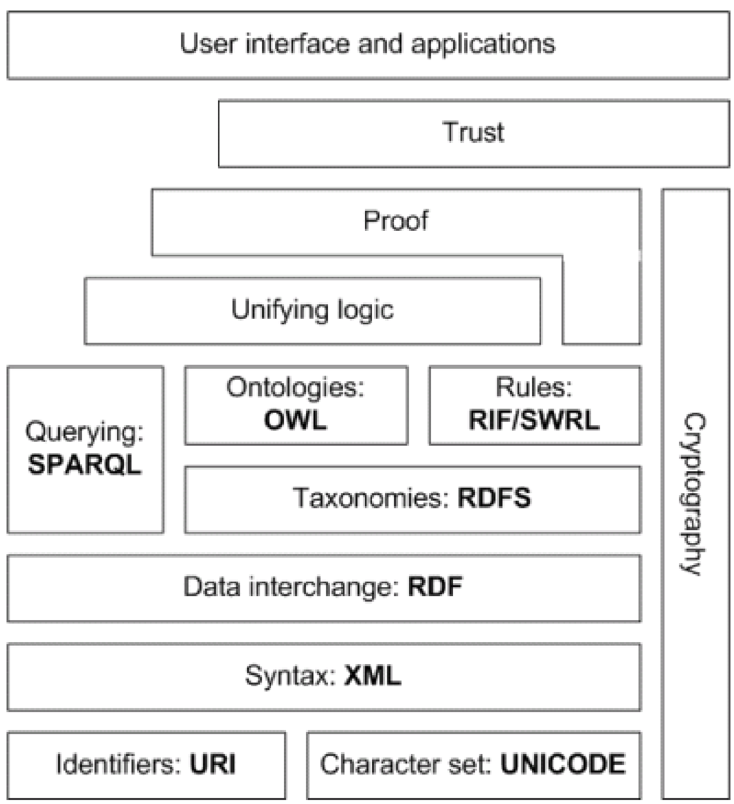


Figure 1 - The Semantic Web Stack

For the Semantic Web to effectively function, the information must be in a structured form and available in a machine-friendly way for computers to perform automated reasoning [1].

Since the goal of the Semantic Web is to link data, every piece of data on the web is represented by a URI, Uniform Resource Identifier. Then, we must give a structured syntax to the data, which is done with XML. It allows the use of tags as labels, but it does not say much about the meaning of the content. RDF is the language used to express meaning in form of a graph. Each node has a unique URI and is described as a triple, meaning that every data is of type subject/predicate/object. Most of the data processed by machines can be described in this way. However, there may be two databases describing the same thing with two different identifiers. The ontologies, collections of information, are used to solve this problem. Usually, an ontology for the web has a set of inference rules and a taxonomy. For the latter, it uses the RDFS to describe the taxonomies of classes and properties. Detailed ontologies are created with OWL, the Web Ontology Language. OWL is syntactically embedded into RDF and was derived from description logics. For querying, it uses SPARQL, Simple Protocol and RDF Query Language. [1,2,7]. All of the concepts above will be further explained on the nexts sections.

3. Knowledge Representation Technologies

As described by Pool [9], “Knowledge is the information about a domain that can be used to solve problems in that domain.”. Knowledge Representation is an area of study of Artificial Intelligence and provides structured data and a set of inference rules [2]. This section describes the languages used to represent knowledge.

3.1 XML

The eXtensible Markup Language (XML) is used to extend the structure of a document adding markup tags surrounding a part of the content [6]. It is usually used to store meta data and has a human understandable form.

It is not a knowledge representation language, but many languages that are used in knowledge representation use the XML syntax. There is also a XMLS, eXtensible Markup Language Schema, which is a template and validation document that is used to define attributes and elements of the XML [2].

The most basic component of the XML is the *element,* which is the text limited by the tags “<> … </>” such as name, description and id. It is also possible to associate attributes to elements.

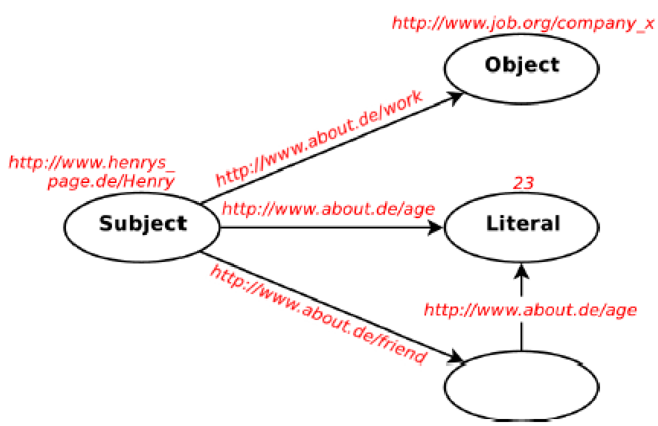
```
<CATALOG>
 <PLANT>
  <COMMON>Bloodroot</COMMON>
  <BOTANICAL>Sanguinaria canadensis</BOTANICAL>
  <ZONE>4</ZONE>
```

```
    <LIGHT>Mostly Shady</LIGHT>
  </PLANT>
</CATALOG>
```

Example 1 - An example of a XML code

3.2 RDF and RDFS

The Resource Description Framework. The data modeling language recommended by the W3C, World Wide Web Consortium for the Semantic Web. All Semantic Web information is stored and represented in the RDF. It aims to serve as a method for description or modeling of information. RDF is designed to be understood by machines and uses XML syntax. It can be represented as a graph which each node is described as a triple. The triple is defined as subject, predicate and object and every data is represented with these three informations.



Example 2 - An example of a RDF graph. From Stroka [2].

```
<?xml version ="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
          xmlns:ab="http://www.about.com/"
          xml:base="http://www.henrys_page.com">
  <rdf:Description rdf:ID="Henry"
   ab:work="http://www.job.com/"
   ab:age="23">
   <ab:friend rdf:nodeID="s3fo" />
  </rdf:Description>

  <rdf:Description rdf:nodeID="s3fo"
   ab:age="23">
  </rdf:Description>
```

```
</rdf:RDF>
```

Example 3 - An example of a RDF/XML code. From Stroka [2].

The Resource Description Language Schema, RDFS extends the vocabulary language that is provided by RDF in order to define terms they intend to use in their document. It is similar to the design in object oriented programming languages. There are four important RDFS vocabulary definitions that are used to define which nodes are connected through a certain property: rdfs:Class, rdf:Property, rdfs:domain and rdfs:range.

3.3 OWL

The Web Ontology Language, OWL, is the schema language of the Semantic Web. OWL enables you to define concepts composably so that these concepts can be reused as much and as often as possible. Composability means that each concept is carefully defined so that it can be selected and assembled in various combinations with other concepts as needed for many different applications and purposes [4]. OWL has three sublanguages: OWL Lite, OWL DL and OWL Full.

OWL Lite supports those users primarily needing a classification hierarchy and simple constraints. OWL DL supports those users who want the maximum expressiveness while retaining computational completeness (all conclusions are guaranteed to be computable) and decidability (all computations will finish in finite time) and OWL Full is meant for users who want maximum expressiveness and the syntactic freedom of RDF with no computational guarantees, as described by McGuinness [10].

An OWL document can contain the following: Header, Classes, Complex Classes Individuals, Properties, Ontology mapping.

As Stroka describes an OWL document [2], in the header it can be found the ontology definition, prior versions, comments and other annotations. The classes can be declared as in object oriented programming to describe real world items. Complex classes can be used to perform set operations, as enumerated and disjoint classes. Instances of the classes are called individuals. There are two types of properties: Object property and Data-type property. The former defines the aggregation between two classes and the latter defines the relation between a class and a literal.

```
Class(pp:animal partial
     restriction(pp:eats someValuesFrom(owl:Thing)))
Class(pp:person partial pp:animal)
Class(pp:man complete
     intersectionOf(pp:person pp:male pp:adult))
Class(pp:animal+lover complete
     intersectionOf(pp:person
            restriction(pp:has_pet minCardinality(3))))
```

Exemple 4 - An example of a OWL class

```
ObjectProperty(pp:eaten_by)
ObjectProperty(pp:eats inverseOf(pp:eaten_by)
                       domain(pp:animal))
ObjectProperty(pp:has_pet domain(pp:person)
                       range(pp:animal))
ObjectProperty(pp:is_pet_of inverseOf(pp:has_pet))
DataProperty(pp:service_number range(xsd:integer))
```

Exemple 5 - An example of a OWL property

```
Individual(pp:Tom type(owl:Thing))
Individual(pp:Dewey type(pp:duck))
Individual(pp:Rex type(pp:dog) value(pp:is_pet_of pp:Mick))
Individual(pp:Mick type(pp:male)
    value(pp:reads pp:Daily+Mirror)
Individual(pp:The42 type(pp:bus)
    value(pp:service_number "42"xsd:integer))
```

Exemple 6 - An example of a OWL individual

4. Query Languages

Query Languages can be, in general, divided into RDF-based and OWL DL-based query languages. They are used to request data from the repositories [2]. Since SPARQL is the language recommended by the W3C, it will be the only query language discussed in this paper.

4.1 SPARQL

The Simple Protocol and RDF Query Language, SPARQL is a RDF-based language for querying RDF data. It enables searches that consists of triple patterns, since it is how RDF data is stored. It also allows queries with disjunctions and conjunctions.

There are four different queries for different goals: SELECT, CONSTRUCT, ASK and DESCRIBE. The first one extracts raw values in a table format. The second extracts information and transform the result into valid RDF. The third is used to answer a simple false/true query. The last one is used to extract an RDF graph [11].

```
PREFIX foaf:  <http://xmlns.com/foaf/0.1/>
SELECT *
WHERE {
   ?person foaf:name ?name .
   ?person foaf:mbox ?email .
}
```

Exemple 7 - An example of a SPARQL search

5. Ontology

Ontologies are collections of information about an specific domain. They are created of knowledge representation and logics and they are necessary for the Semantic Web in interest of enhancing the meaning about the data and formalising it. Formalising the knowledge allows it to connect to other formalised knowledge, which is fundamental for the Semantic Web [13].

Description logics is a family of languages of knowledge representation. It is used for formal reasoning in AI on an application domain. It provides the logical formalism for ontologies that the Semantic Web needs. The language OWL is based on description logics and expressible in RDF. It allows the development of ontologies more expressible than RDFS. Additionally, it allows possible efficient subsumption inferences and classification.

6. Rules

Rules are used to perform proof tests without a full logic machinery. They are also used to create a set of conditions on top of dynamic knowledge captured that must be accomplished in order to achieve the results of that specific rule.

On the Semantic Web Stack, the SWRL, Semantic Web Rule Language, is the technology used. It covers from derivation and transformation rules to reaction rules, permitting it to specifying queries and inferences in ontologies, mapping between ontologies and dynamic web behaviours of workflows, services and agents, as described by González [13] .

7. Logic

The logic layer provides features of FOL or First Order Logic, which is a formal system that uses quantified variables over (non-logical) objects [15]. This makes the Semantic Web abled to use all the capabilities of logic available at a reasonable computation cost.

8. Proof

Semantic Web makes use of inference engines. An inference engine interprets and evaluates the facts in the knowledge base in order to provide an answer and also, when asked why it arrived to a conclusion, it provides proof to its results. Furthermore, inference engines problems are open questions that may require great or infinite answer time, which can be solved by adding proofs in a way that when the problem is faced, the proofs are written making it easier to solve due to the more constrained reasoning context.

9. Trust

On the top layer of the Semantic Web Stack is trust. This is where agents will work on a full-featured Semantic Web. It uses the features of the layers below, however, it does not provide the functionality to trustily bind statements with their responsible parts. This is achieved with supplementary technologies.

Using digital signature and encryption, the trust web will extensively use public keys, as it is done already. This web of trust is based on the graph structure of web. It has reasoning engines that complements with digital signatures to create trust-engines.[13].

10. Agents

The focus of the Semantic Web is to enable machines to process information from many resources in order to become a support for the end users, performing sophisticated tasks on their behalf. To perform such tasks, intelligent agents will be used. As Heylighen explained in [14], “An *agent* is a (typically small) program or script, which can travel to different places and make decisions autonomously, while representing the interests of its user.”

Agents must have an understanding of the meaning of the content available and also must be able to communicate with other agents in order to locate meaningful resources on the web. They must combine resources, recognize, interpret and respond to communication acts from other agents. Agents interact with other agents from different places, which does not allow the assumption that the terms used between them will be the same. Consequently, the agents will need to publicly declare the terms they are using and their meaning, resulting in an agent’s ontology.

Agent ontologies define the resources that are being processed in a metadata. Terms in different ontologies are interconnected through the World Wide Web. Thus, when two agents meet, they can use connections between their ontologies to understand one another [13].

11. Challenges

The Semantic Web face a number of challenges to be fully implemented. Automated reasoning systems will have to solve all this issues in order to propitiate its creation.

Among all the challenges there are the problem with the extreme amount of data on the World Wide Web. Automated reasoning systems will deal with huge inputs. The natural language vagueness is also a problem, e.g. imprecise descriptions as “tall” or “young”. This is usually dealt with by using fuzzy logic. There is the uncertainty issue in concepts with uncertain values, like the symptoms of a patient. Uncertainty is generally dealt with probabilistic reasoning techniques. There also may be an inconsistency on the creation of ontologies which makes deductive reasoning techniques fails. At last, but not least, the deceit problem is real on the internet, when the author of the information is intentionally feeding the base of data with misleading contents.

Theses challenges are meant to be treated on the logic and proof layers of the Semantic Web Stack [16].

11. Final Considerations and Personal Opinion

The whole concept of the Semantic Web presented by Tim Berners-Lee is still being developed and, as presented on the previous section, faces several challenges to be fully implemented. However, a lot of features and languages to support their objectives are already implemented and long being developed in order to achieve the so called web of data.

Based on my research, from all the papers and presentations that I have read, there are a big number of groups and institutions committed to the development of the Semantic Web. Although when compared to the total number of data currently available on the internet it may seem not much, the amount of semantic data currently available is considerably big and it is continuously growing. This, indoubtably, represents the future of the web. Many concepts can already be seen today, on simple google searches, as "*How is the weather like today?*", which results on the actual answer instead of a number of links to weather websites. This only shows that we are currently witnessing what is considered by many the biggest changing on the World Wide Web since its creation.

Concluding, with this project I had the opportunity to better understand the concepts of a big field of study which I plan to continue studying. My intentions from the end of this project and on are now searching for what has already been done on the domain of information of law concepts, which I have already worked with, meaning all the semantic data already collected and converted to RDF up until this moment and develop a system or a web service that can contribute to the spread and usage of this information.

References

[1] Berners-Lee, Tim; James Hendler; Ora Lassila, "The Semantic Web". Scientific American Magazine, May 17, 2001. Retrieved on November 2, 2014;

[2] Stroka, Stephanie, "Knowledge Representation Technologies in the Semantic Web". Salzburg University of Applied Sciences. Retrieved on November 2, 2014

[3] World Wide Web Consortium (W3C), "W3C Semantic Web Activity". November 7, 2011. Retrieved on November 2, 2014;

[4] Gonzales, Rob. "Introduction to the Semantic Web". Cambridge Semantics. http://www.cambridgesemantics.com/semantic-university/introduction-to-the-semantic-web Retrieved on November 4, 2014;

[5] Hendler, James. "Agents and the Semantic Web". University of Maryland. http://cs.rpi.edu//~hendler/AgentWeb.html Retrieved on November 6, 2014;

[6] Bray, Tim, et al. "Extensible markup language (XML)." World Wide Web Consortium Recommendation REC-xml-19980210. (1998). Retrieved on November 20, 2014;

[7] Ross, Greg. “An introduction to Tim Berners-Lee's Semantic Web”. January 31, 2005. http://www.techrepublic.com/article/an-introduction-to-tim-berners-lees-semantic-web/ Retrieved on November 20, 2014;

[8] Obitko, Marek. “Introduction to Ontologies and Semantic Web”. 2007. http://obitko.com/tutorials/ontologies-semantic-web. Retrieved on November 24, 2014;

[9] Poole, David L.; Mackworth, Alan K. “Artificial Intelligence: foundations of computational agents”. Cambridge University Press, 2010. Retrieved on November 24, 2014;

[10] McGuinness, Deborah L.; Harmelen, Frank V. "OWL web ontology language overview." W3C recommendation 10.10 (2004): 2004. Retrieved on November 24, 2014;

[11] Prud, Eric, and Andy Seaborne. "Sparql query language for rdf." (2006). W3C recommendation 15.1 (2008): 2008. Retrieved on November 24, 2014;

[12] Bechhofer, Sean, et al. “Tutorial on OWL”. 2003. http://www.cs.man.ac.uk/~horrocks/ISWC2003/Tutorial/examples.pdf.Retrieved on November 24, 2014;

[13] González, Roberto G. "A semantic web approach to digital rights management.". Department of Technologies. Universitat Pompeu Fabra. November, 2005.;

[14] Heylighen, F. & Bollen, J.: "The World-Wide Web as a Super-Brain: from metaphor to model". "Cybernetics and Systems '96". Austrian Society for Cybernetics, pp. 917-922, 1996;

[15] Wikipedia, “First-order Logic” http://en.wikipedia.org/wiki/First-order_logic. Retrieved on November 26, 2014.

[16] Wikipedia, “Semantic Web” http://en.wikipedia.org/wiki/Semantic_Web. Retrieved on November 26, 2014.