Oct 26th, 11:00 AM - 12:30 PM

Semantic Web Technologies

Dukagjin Hyseni
dukagjinhyseni@hotmail.com

Semantic Web Technologies

Dukagjin Hyseni

UBT – Higher Education Institution, Lagjja Kalabria, 10000 p.n., Prishtine, Kosovo

Abstract. This study attempts to highlight the great importance of developing Semantic Web as one of the best discovery of better data management and presentation within the WWW. Since the W3C's was discovered, initially providing classic web content as web 1.0 that had link / hyperlink of document`s location, then web 2.0 as web-applications have more advanced technologies to connect data, and finally semantic web as extension of web 3.0 also known as Linked Data.

The results show that in addition to the rapid development of the Semantic Web, the demand to use its features by data publishers and data readers is rapidly expanding due to the time saving to publish multiple times the same data on other web pages.

Moreover, we will present the features of the Semantic Web, its technologies, development history, advantages and weaknesses, the potential benefits, and so on, including standards, frameworks, and programming languages that are being used in its development like: RDF (Resource Description Framework), XML etc.

Key words: Semantic web, Linked Data, W3C, RDF, XML.

Introduction

Since Tim Berners-Lee's original idea for a global system of interlinked hypertext documents from 1989, the World Wide Web has grown into the world's biggest pool of human knowledge. Over the past few years, the Web has changed the way people communicate and exchange information. In 1998, the size of the Web was estimated to exceed 300 million pages with a growth rate of about

20 million per month. The real size of the Web today is difficult to measure, although Web search indices cite a lower band number of unique and meaningful Web pages. The explosion of Web documents and services would not be so critical if users could easily retrieve and combine the information needed. Since Web documents are at best semi-structured in simple natural language text, they are vulnerable to obstacles that prevent efficient content retrieval and aggregation. An increasing problem is the number of languages used on the Web.

The Semantic Web term was popularized by Tim Berners-Lee and later elaborated in 2001. The first part of his vision for the Semantic Web was to turn the Web into a truly collaborative medium—to help people share information and services and make it easier to aggregate data from different sources and different formats. The second part of his vision was to create a Web that would be understandable and processable by machines. The key to machine-processable data is to make the data smarter. Smart data continuum of Semantic Web is consist from: Text and Databases, XML documents for single domains, Taxonomies, Ontologies and automated reasoning.

Semantic Web refers to the W3C`s vision of the Web of linked data. Semantic Web technologies enable people to create data stores on the Web, build vocabularies, and write rules for handling data. Linked data are empowered by technologies such as RDF, SPARQL, OWL and SKOS.

Semantic web technologies

Currently, the World Wide Web is primarily composed of documents written in HTML (Hyper Text Markup Language), a language that is useful for publishing information. HTML is a set of "markup" symbols contained in a Web page intended for display on a Web browser. During the first decade of its existence, most of the information on the Web is designed only for humanconsumption. Humans can read Web pages and understand them, but their inherent meaning is not shown in a way that allows their interpretation by computers.
The information on the Web can be defined in a way that it can be used by computers not only for display purposes, but also for interoperability and integration between systems and applications. One way to enable machine-to-machine exchange and automated processing is to provide the information in such a way that computers can understand it. This is precisely the objective of the semantic Web – to make possible the processing of Web information by computers.
"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." (Berners-Lee, Hendler et al. 2001). "The next generation of the Web will combine existing Web technologies with knowledge representation formalisms". (Grau, 2004)
The third common use of the term Semantic Web is to identify a set of technologies, tools and standards which form the basic building blocks of a system that could support the vision of a Web imbued with meaning. The Semantic Web has been developing a layered architecture, which is often represented using a diagram first proposed by Tim Berners-Lee, with many variations since.

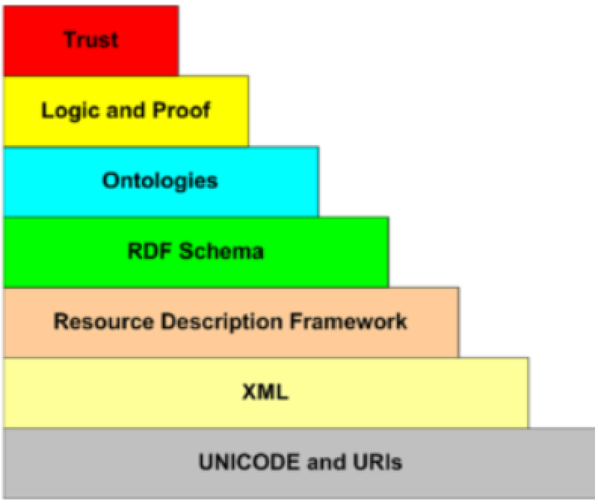

Figure 1: Semantic Web layered architecture

Resource Description Framework (RDF)

The Semantic Web aims to build a common framework that allows data to be shared and reused across applications, enterprises, and community boundaries. It proposes to use RDF as a flexible data model and use ontology to represent data semantics. Currently, relational models and XML tree models are widely used to represent structured and semi-structured data. But they offer limited means to capture the semantics of data. RDFS and OWL ontologies can effectively capture data semantics and enable semantic query and matching, as well as efficient data integration.

The Resource Description Framework (RDF) – now around for over a decade already as well – is the basic data model for the Semantic Web. It is built upon one of the simplest structures for representing data: a directed labeled graph. An RDF graph is described by a set of triples of the form (Subject Predicate Object), also called statements, which represent the edges of this graph. RDF's flat graph-like representation has the advantage of abstracting away from the data schema, and thus promises to allow for easier integration than customized XML data in different XML dialects: whereas the integration of different XML languages requires the transformation between different tree structures using transformation languages such as XSLT or XQuery. While the normative syntax to exchange RDF, RDF/XML, is an XML dialect itself, there are various other serialization formats for RDF, such as RDFa, a format that allows to embed RDF within (X) HTML, or non-XML representations such as the more readable Turtle syntax; likewise RDF stores (e.g. YARS2) normally use their own, proprietary internal representations of triples, that do not relate to XML at all.4

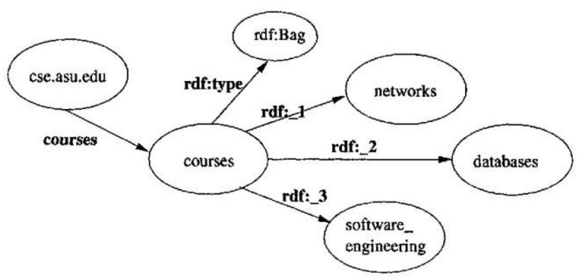

Figure 2. Graph model of RDF/XML implementation

SPARQL

SPARQL a query language for RDF is used to return and manipulate data from databases that are stored in Resource Description Format. It is the recommendation of (DAWG) Data Access Working Group on RDF under World Wide Web Consortium, and is recognized as one of the key technologies of the semantic web.

The SPARQL Protocol and RDF Query Language (SPARQL) is the de-facto standard used to query RDF data. While RDF and the RDF Schema provide a model for representing Semantic Web data and for structuring semantic data using simple hierarchies of classes and properties, respectively, the SPARQL language and protocol provide the means to express queries and retrieve information from across diverse Semantic Web data sources. The SPARQL query language was developed for the RDF layer of the Semantic Web architecture (see Fig.6.1). The query language has been developed without considering the other core languages of the Semantic Web, namely RDFS, OWL and RIF. The SPARQL is a matching graph pattern language. It defines a set of graph patterns, the simplest being the triple pattern. A triple pattern is like a normal RDF triple but with the possibility of a variable instead of an RDF term in the subject, predicate, or object positions. SPARQL introduces the notion of variable binding, that is, a pair (variable, RDF term), where variable is a query variable of interest indicated by ? or $ and the RDF term is the value assigned to the variable after the query has been executed. Similar to the namespace mechanism used for writing RDF/XML, SPARQL allows the definition of prefixes for namespaces. Prefixes are used inside a query to increase its readability. SPARQL introduced a set of constructs and clauses that could be part of a query.

The SELECT clause identifies the variables to appear in the query results and the WHERE clause provides the basic graph pattern to match against the data graph.

```
PREFIX ex: <http://example.org/> .
SELECT ?name
WHERE
{
ex:john, ex:hasName, ?name .
}
```

Figure 3 – Simple SPARQL query

Web Ontology Language (OWL)

OWL stands for Web Ontology Language and was founded on 2004. OWL is such a language for modeling such complex knowledge, expressive representation languages based on formal logic are commonly used. This also allows us to do logical reasoning on the knowledge, and thereby enables the access to knowledge which is only implicitly modeled. Since 2004 OWL is a W3C recommended standard for the modeling of ontologies, and since then has seen a steeply rising increase in popularity in many application do-mains. Central for the design of OWL was to find a reasonable balance be-tween expressivity of the language on the one hand, and efficient reasoning, i.e. scalability, on the other hand. This was in order to deal with the general observation that complex language constructs for representing implicit knowledge usually yield high computational complexities or even undecidability of reasoning, and therefore unfavorable scalability properties. OWL has three sublanguages: OWL Full, OWL DL and OWL Lite. OWL documents are used for modeling OWL ontologies. Two different syntaxes have been standardized in order to express these. One of them is based on RDF and is usually used for data exchange. It is also called OWL RDF syntax since OWL documents in RDF syntax are also valid RDF documents. The basic building blocks of OWL are classes and properties, which we already know from RDF(S), and individuals, which are declared as RDF in-stances of classes. OWL properties are also called roles, and we will use both notions interchangeably.

```
<rdf:RDF
   xmlns:owl ="http://www.w3.org/2002/07/owl#"
   xmlns:rdf ="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
   xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
   xmlns:xsd ="http://www.w3.org/2001/XMLSchema#">
   <owl:Ontology rdf:about="">
        <rdfs:comment>An example OWL ontology</rdfs:comment>
        <owl:imports rdf:resource="http://www.mydomain.org/persons"/>
        <rdfs:label>University Ontology</rdfs:label>
   </owl:Ontology>
   <owl:Class rdf:ID="academicStaffMember"></owl:Class>
   <owl:Class rdf:ID="associateProfessor">
        <rdfs:subClassOf rdf:resource="#academicStaffMember"/>
    </owl:Class>
    …
</rdf:RDF>
```

Figure 4 – OWL Example

The OWL language provides mechanisms for creating all the components of an ontology: concepts, instances, properties (or relations) and axioms. Two sorts of properties can be defined: object properties and data type properties. Object properties relate instances to instances. Data type properties relate instances to data type values, for example text strings or numbers. Concepts can have super and sub concepts, thus providing a mechanism for subsumption reasoning and inheritance of properties.

Development and usability of Web Semantic

Since time that Lee has developed semantic web, there are several version of development of semantic web. Thirty years after the birth of the Web, to have a Web linking applications, things, people, data, etc. Tim Berners-Lee insisted very early on the need to provide on the Web "more machine oriented semantic information, allowing more sophisticated processing" (Berners-Lee et al., 1994). To bootstrap that evolution Tim Berners-Lee then proposed in September 1998 a "Semantic Web Road map" (Berners-Lee, 1998) giving 20 years ago the blue prints of the architecture of the Semantic Web. In 1999 the first versions of RDF and RDFS were published by the W3C and the vision of a Semantic Web was then made visible to a broad audience in 2001 with an article in the Scientific American (Berners-Leet al., 2001). This well-known article presents the Semantic Web as an extension of the existing document-based Web with a Web of structured data and formal semantics better enabling computers and people to work in cooperation. A few years later, Tim Berners-Lee will be again instrumental in pushing what can be seen as a first wave of deployment of the Semantic Web with the Linked Data principles and the Linked Open Data 5-star rules (Berners-Lee, 2006) leading to the publication and growth of linked open datasets weaving a Web of Linked Data. It all started with the first international Semantic Web Working Symposium (SWWS), a workshop held in Stanford, Palo Alto, the 30th of July and 1st of August 2001. The following year the symposium became the International Semantic Web Conference (ISWC) series. Nowadays, Semantic Web not only has its conferences (e.g. ISWC, ESWC, SemTech, SemWeb.Pro) and journals (e.g. Semantic Web Journal, Journal of Web Semantics) but is also an established topic of older conferences and journals from other domains (e.g. The Web Conference WWW, VLDB, EKAW, IJCAI/ECAI, WI, etc.). 7
Semantic Web from year to year is developing rapidly thanks to the development of various platforms for designing and building web semantic technologies. Based on this technological development, usability of web semantic is increasing rapidly. This usability we can verify also by the huge number of Facebook users with 2.45 billion, Google averages over 2 trillion Google searches a day per year 2019 or over 63,000 search queries done per second and some others web- sites that are example and type of web semantics or linked data on web.8 Regarding to the data of usage of web semantics, it seems that usage of web-semantics is too high so we are going to present some graphs of using web semantics.

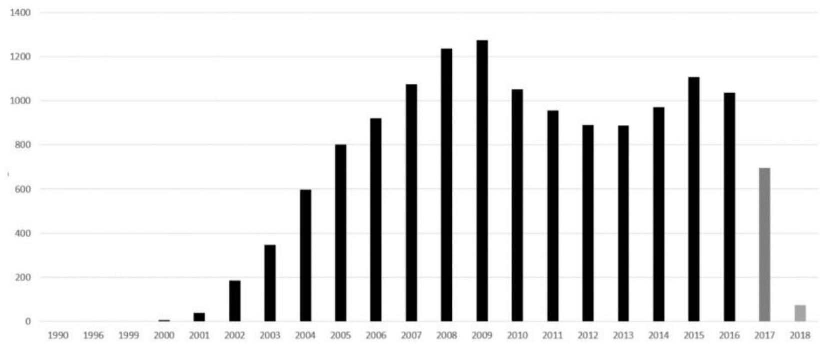

Figure 5 - Yearly distribution of 14157 documents found on the Web of Science on the topic "Semantic Web", note

that at the moment of the survey the years 2017 and 2018 were largely incomplete

Figure 6 - Distribution per country of 14157 documents found on the Web of Science on the topic "Semantic Web"

Conclusion

This paper has primarily described how RDF data store need to be represented as a schema in which the prepositions will be used as properties. The properties will be then interpreted by the SPARQL engines. The property based schema can extend to OWL ontologies which are RDF serialized.

The SPARQL engine will however only return search based on preposition when the underlying data set is got the appropriate schema. Thus all data sets of the future will need to be RDF repositories. It is therefore a subject of research to have software agents which can convert Non- RDF datasets to RDF data sets. The agents can a part of the SPARQL engine itself which can interpret Non-RDF datasets to be RDF data sets only for the purpose of query.

REFERENCES

1. Berners-Lee, T., J. Hendler, et al. (2001). The Semantic Web. Scientific American. May 2001
2. Berners-Lee (1998). Semantic Web Road Map. W3C.Available at: http://www.w3.org/DesignIssues/Semantic.html[last accessed 25/12/19]
3. Vijayan Sugumaran, Jon Atle Gulla, Applied Semantic Web Technologies, CRC Press (New York), 4-5
4. Jorge Cardoso & Amit Sheth, The Semantic Web and its Applications, Springer (Berlin, 2006)
5. Dr Brian Matthews & Leigh Dodds, Semantic Web Technologies, CCRLC Rutherford Appleton Laboratory (England, 2005), 4
6. Axel Polleres, Semantic Web Technologies: From Theory to Practice, Technischen Universitat Wien (Galway, 2001)
7. Sneha Kumari & Dr.Rajiv Pandey & Amit Singh & Himanshu Pathak, Sparql: Semantic Information Retrieval by Embedding Prepositions, IJNSA, (Lucknow, India, 2014), 49
8. Dieter Fensel & Federico Michele Facca & Ioan Toma & Elena Simperl, Springer Heidelberg Dordrecht London New York, (2011), 94 – 95
9. Fabien Gandon, A Survey of the First 20 Years of Research on Semantic Web and Linked Data, 10.3166/ISI.23.3-4.11-56ff.ffhal-01935898, (2018) 12
10. Ben Adida, Mark Birbeck, Shane McCarron, and Steven Pemberton. RDFa in XHTML: Syntax and Processing. W3C recommendation, W3C, October 2008. Available at: http: //www.w3.org/TR/rdfa-syntax/, last accessed 26/12/109
11. Steinmetz, N., Lausen, H., Brunner, M.: Web service search on a large scale. In: Proceedings of the 7th International Joint Conference on Service Oriented Computing (ICSOC 2009)