Semantic Web: vision and reality

Mile Jovanov, Marjan Gusev

Institute of Informatics, FNSM, Gazi Baba b.b., 1000 Skopje
{mile, marjan}@ii.edu.mk

Abstract. Semantic Web is set of technologies currently developing. This paper presents an overview of the technology already developed, and part that should be built in the future. FOAF, as example of decentralized SW technology is explained. In addition, the circumstances in which SW is developing, and chances for the success of the idea are discussed.

1 Introduction

Nowadays, web sites and other information systems have predetermined how information is presented, how information is used (combined, manipulated) and sometimes what things are automated. This makes it hard to provide access using a non-standard device. It is very hard to automate the task of information combining, also. Automation of unforeseen situations is impossible.

Web with its content was built for humans, so human interpretation is needed to use information and accomplish some tasks on the Web. Content interpretable by the machines is needed. Information with accessible formal semantics could allow machines to reason.

Semantic Web is an attempt to address the initial goal of the web enabling automation. Short term goal is interoperability, and long term goal is to make computers work on our behalf instead of using them like tools [1].

An example of the short term goal is the possibility someone's web application about social networking to use the data about people entered in all other open social networking applications (done with SW technologies).Using it, meaning displaying it, searching thru it, and some similar tasks.

An instance of the long term goal is, for example, following situation: The computer of student A, on the demand of the student, finds a way to influence professor's B decision about the grade of the student.

2 Semantic Web

During the years of evolution, several definitions about semantic web can be found. According to the founders of the idea: 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 [2]. According to **W3C** (World Wide Web Consortium): The Semantic Web is the abstract representation of data on the World Wide Web, based on the RDF standards and other standards to be defined. It is being developed by the W3C, in collaboration with a large number of researchers and industrial partners [3]. The latest attempt to define Semantic Web that can be found at http://www.w3.org/2001/sw/ is: The Semantic Web is about two things. It is about common formats for integration and combination of data drawn from diverse sources, where on the original Web mainly concentrated on the interchange of documents. It is also about language for recording how the data relates to real world objects. That allows a person, or a machine, to start off in one database, and then move through an unending set of databases which are connected not by wires but by being about the same thing.

Goals of Semantic Web are to develop enabling standards and technologies designed to help machines understand more information on the Web so that they can support richer discovery, data integration, navigation, and automation of tasks. This means that one should expect to receive more exact result when searching for information. Machines should know when to integrate information from different sources and what information to compare.

"The layer cake" that represents the layers in Semantic Web has evolved and it's still changing (Fig. 1(a), Fig. 1(b)). This shows that semantic web standards and technologies are work in progress and new ideas for better "infrastructure" and faster adoption are given continuously.

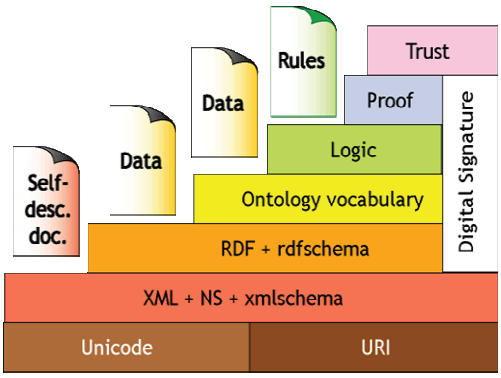

Fig. 1(a). SW Layer Cake, 2001
http://www.w3.org/2001/09/06-ecdl/slide17-0.html

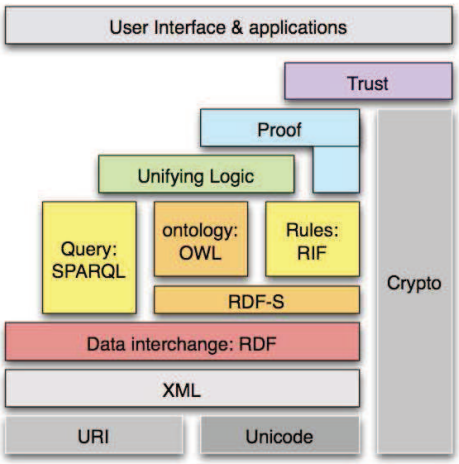

Fig. 1(b). SW Layer Cake, 2006
http://thefigtrees.net/lee/blog/2006/11/semantic_web_technologies_in_t.html

First ideas about Semantic Web were given at the beginning of 21 century. Starting 1996 W3C consortium have worked on developing standards like XML (1996-2004), RDF (1997-2004), RDFS (1998-2004), OWL (2002-2004), SPARQL (2004-). Most of them have already reached stage of recommendation.

3 Semantic web technologies

Main goal of Semantic Web is inclusion of metadata in web documents. Metadata, or a structured data about data, will allow machines to interpret information more easily. Of course, descriptive information about an object or resource has to be well structured in order interoperability to be possible. At the first level, digital library community have done a lot in definition of sets of metadata. Prominent examples are Dublin Core, XMP, IPTC, etc. Dublin Core metadata are developed by Dublin Core Metadata Initiative [4], an organization providing an open forum for the development of interoperable online metadata standards that support a broad range of purposes and business models.

Semantic Web Technologies are the foundations of a systematic approach to creating "smart data". Starting with text documents and database records, continuing with XML documents that use single vocabularies, and then with XML taxonomies and documents with mixed vocabularies, technologies come to level of XML ontology and automated reasoning.

3.1 XML into Semantic Web

XML (eXtensible Markup Language) creates application-independent documents and data. It has a standard syntax for metadata and also, a standard structure for both

documents and data. XML practically represents the *syntactical foundation layer* of Semantic Web.

There are, also, limitations of XML like the possibility to say the same thing in many ways because multiple valid structures for the same data can exist. XML isn't forcing a common interpretation of a data. Different words can be used for the same concept: price vs. cost or address vs. location, for example

3.2 RDF

RDF (Resource Description Framework) is an infrastructure that enables the encoding, exchange and reuse of structured metadata. [5] It is a foundation for processing metadata, providing ***interoperability*** between applications that exchange machine-understandable information on the Web. RDF is an application of XML that imposes needed structural constraints to provide unambiguous methods of expressing semantics.

The model that can be represented by RDF consists of resources, properties and statements.

Resources are things being described by RDF expressions. A resource may be an entire Web page; may be a part of a Web page; may be a whole collection of pages. The resource not necessarily exists on the web. It could be an object from the real world.

Properties represent a specific aspect, characteristic, attribute, or relation used to describe a resource. Each property has a specific meaning that defines its permitted values, the types of resources it can describe, and its relationship with other properties.

A specific resource together with a named property plus the value of that property for that resource is an RDF statement.

RDF Graph Diagram of Statement is given in Fig. 2.

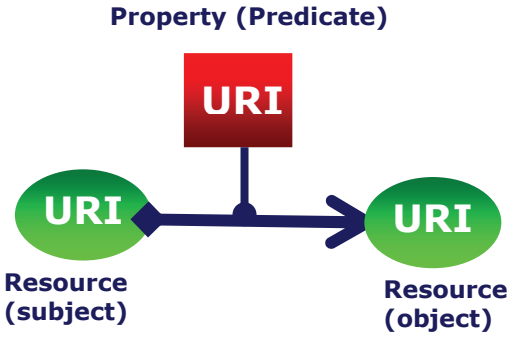

Fig. 2. Statement in RDF, graph diagram

3.3 XML vs. RDF

Because XML was designed for documents, not for data many features (like attributes and entities) are document-oriented, and not for expressing data. There are many ways to say the same thing in XML, hybrid tree structure, confusing and non standard can be formed. Even basic operations like merging are complex. On the other hand, RDF was designed for statements, or data. The number of changes that can be made to a triple is fairly small. Structures are simple, they are triples. Merging two documents is simple combination two into one.

3.4 Ontology for Semantic Web

A program that wants to compare or combine information across the two databases has to know that two terms are being used to mean the same thing. Ontologies can play a crucial role in enabling Web-based knowledge processing, sharing, and reuse between applications. [6]

According to Tom Gruber, ontology defines a set of representational primitives with which to model a domain of knowledge or discourse. The representational primitives are typically classes (or sets), attributes (or properties), and relationships (or relations among class members). [7]

The most typical kind of ontology for the Web has *taxonomy* and a set of *inference rules*.

The taxonomy defines classes of objects and relations among them. Classes, subclasses and relations among entities are a very powerful tool for Web use.

Using inference rules a program can deduce new instances. The computer doesn't truly understand any of this information, but it can now manipulate the terms much more effectively in ways that are useful and meaningful to the human user.

Many ontology languages like **XOL** (XML-based Ontology Exchange Language), **SHOE** (Simple HTML Ontology Extension), **OML** (Ontology Markup Language), **RDF(S)** (Resource Description Framework (Schema)), **DAML+OIL** (DARPA Agent Markup Language + OIL), **OWL** (Web Ontology Language) exist.

OWL is the ontology language developed and adopted by W3C as a standard semantic web technology. It is a semantic markup language for Web resources. It is built on earlier W3C standards such as RDF and RDF Schema, and extends these languages with richer modeling primitives. Provides modeling primitives commonly found in frame-based languages.

3.5 Logic, Proof, Trust

Parts of the Semantic Web that are into phase of development are logic, proof and trust. Logic should allow the computer to reason (by inference) using all stated logical principles. Proof will be possible after some systems that follow logic are built, so they can be used to prove things.

Other important issue is whether one can trust RDF statements. It's the issue of trust.

4 Example SW project: FOAF

FOAF (http://www.foaf-project.org/), like the Web itself, is a linked information system. It is built using decentralized Semantic Web technology, and has been designed to allow for integration of data across a variety of applications, Web sites and services, and software systems. To achieve this, FOAF takes a liberal approach to data exchange. It does not require one to say anything at all about him or others, nor does it place any limits on the things that can be said or the variety of Semantic Web vocabularies that can be used in doing so.

The FOAF project is based around the use of *machine readable* Web homepages for people, groups, companies and other kinds of thing. "FOAF vocabulary" (Fig. 3) provides a collection of basic terms that can be used in these Web pages. At the heart of the FOAF project is a set of definitions designed to serve as a dictionary of terms that can be used to express claims about the world.

Initial focus is the description of people, since people are the things that link together most of the other kinds of things described in the Web. People make documents. They attend meetings, are depicted in photos, and so on.

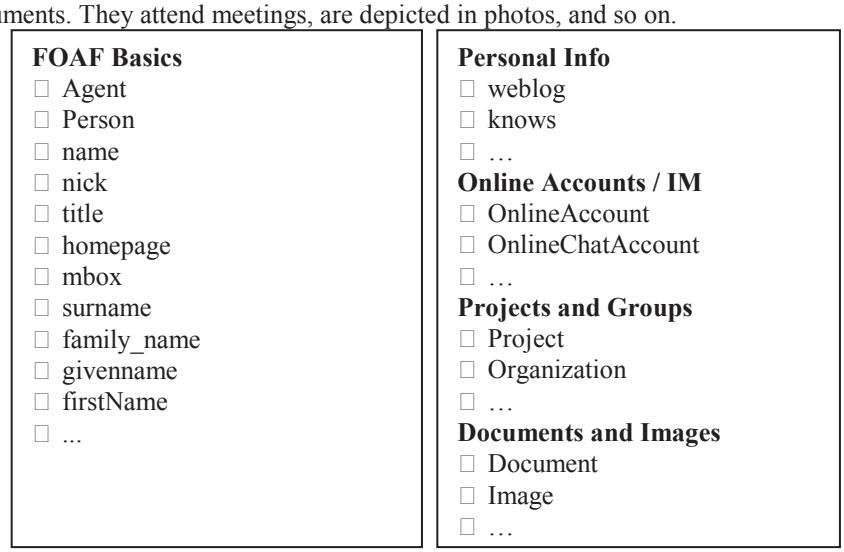

Fig. 3 Part of FOAF terms, grouped in broad categories

The basic idea is pretty simple. If people publish information in the FOAF document format, machines will be able to make use of that information. If those files contain "see also" references to other such documents in the Web, a machine-friendly version of today's hypertext Web will be produced. Computer programs will be able to surf around a Web of documents designed for machines rather than humans, storing the information they find, keeping a list of "see also" pointers to other documents,

checking digital signatures (for the security minded) and building Web pages and question-answering services based on the harvested documents.

5 Odds of SW long term goal

Based on current development of technology, one could assume that the long term goal of Semantic Web is reachable. It is important to say that the goal would not be reached as fast as it was predicted but the creators of the idea for SW. In the time of promoting the idea, probably the authors were aware of that, but the buzz had to be created. This kind of buzz is always useful and it allows some new technology to be quickly developed and adopted.

The idea has lost a breath in the moment when it was clear that the results would not come quickly. A new challenge for the promoters of the idea was the set of applications that allowed the social aspect of the web. This set of already present applications was called ***Web2.0***.

There was a lot of debates and argument saying that there is no qualitative jump in the technology. At the end a new reality – accepted term Web2.0 – was recognized. At that moment, Semantic Web was named Web3.0, and the "short term goal of SW – interoperability" was promoted.

An effort will be made the term, Web3.0, to be accepted as well as Web2.0. But the challenge is huge, because there could be a moment when someone will use the slow development of SW infrastructure and will declare some already present technology as Web4.0 (now when the race with numbers, i.e. versions is already started). Will that be a challenge of putting the Semantic Web technology into history, as some "unsuccessful version of the web"?

Anyhow, in the light of the instance of the long term goal given in the introduction of this paper, one could say that if the infrastructure of SW is consistently implemented, the computer (actually the agent installed on it) could look thru the social networks and sets of RDF statements about the student and the teacher and could resolve which persons "link" them (an almost certainly there will be some). A FOAF could be just one example of RDF statement set.

The possibility of choosing the one person with most influence to the teacher, or the most likely link to the teacher, will be also present. (That will be the time when all the teachers will have to grade their students only with A.)

6 Conclusion

Semantic Web is set of technologies currently developing. Here, an overview of the technology already developed was given. The parts of Semantic Web that should be built in the future were pointed out. FOAF, as example of decentralized SW technology was given. In the last part, the circumstances in which SW is developing, and chances for the success of the idea were discussed. It was concluded that if the infra-

structure of SW is consistently implemented, this future web could be a powerful tool in a world to come.

References

1. Ora Lasila, “Towards the sematic web”, W3C Semantic Tour, London, June 2003, http://www.w3c.rl.ac.uk/pastevents/TowardsTheSemanticWeb.pdf
2. Tim Berners-Lee, James Hendler, Ora Lassila, “The Semantic Web”, Scientific American 284(5):34-43, May, 2001. http://www.sciam.com/article.cfm?id=the-semantic-web
3. http://www.digicult.info/pages/resources.php?t=10, quotation from w3.org/2001/sw/
4. Dublin Core Metadata Initiative (DCMI), dublincore.org
5. Eric Miller, “An Introduction to the Resource Description Framework”, D-Lib Magazine, 1998
6. Joongmin Choi, “Semantic Web: Overviews and Trends”
7. http://tomgruber.org/writing/ontology-definition-2007.htm