

UNIVERSITÀ DELLA SVIZZERA ITALIANA

FACULTY OF COMPUTER SCIENCE

Master in Artificial Intelligence

Topic: Semantic Web

By Heider Jeffer

14 December - 20 December 2014

Instructor: Prof. Mehdi Jazayeri

speaker: Prof. Grigoris Anoniou

# Summary

The Semantic Web is an evolving development of the World Wide Web in which the meaning (semantics) of information and services on the web is defined, making it possible for the web to” understand” and satisfy the requests of people and machines to use the web content.

the semantic web comprises a set of design principles, collaborative working groups, and a variety of enabling technologies, some elements of the semantic web are expressed as prospective future possibilities that are yet to be implemented or realized, and other elements of the semantic web are expressed in formal specifications, some of these include Resource Description Framework (RDF), a variety of data interchange formats (e.g. RDF/XML, N3, Turtle, N-Triples), and notations such as RDF Schema (RDFS) and the Web Ontology Language (OWL), all of which are intended to provide a formal description of concepts, terms, and relationships within a given knowledge domain.

Semantic publishing will benefit greatly from the semantic web. In particular, the semantic web is expected to revolutionize scientific publishing, such as real-time publishing and sharing of experimental data on the Internet. Humans are capable of using the Web to carry out tasks such as finding the Finnish word for ” monkey”, reserving a library book, and searching for a low price for a DVD. However, a computer cannot accomplish the same tasks without human direction because web pages are designed to be read by people, not machines. The semantic web is a vision of information that is understandable by computers so that they can perform more of the tedious work involved in finding, sharing, and combining information on the web. Semantic Web application areas are experiencing intensified interest due to the rapid growth in the use of the Web, together with the innovation and renovation of information content technologies.

The Semantic Web is regarded as an integrator across different content and information applications and systems and provides mechanisms for the realization of Enterprise Information Systems. The rapidity of the growth experienced provides the impetus for researchers to focus on the creation and dissemination of innovative Semantic Web technologies, where the envisaged Semantic Web is long overdue. Often the terms Semantics, metadata, ontologies, and Semantic Web are used inconsistently. In particular, these terms are used as everyday terminology by researchers and practitioners, spanning a vast landscape of different fields, technologies, concepts, and application areas. Furthermore, there is confusion about the current status of the enabling technologies envisioned to realize the Semantic Web.

# Limitations of HTML

Many files on a typical computer can be loosely divided into documents and data. Documents like mail messages, reports, and brochures are read by humans. Data, like calendars, address books, playlists, and spreadsheets are presented using an application program that lets them be viewed, searched, and combined in many ways.

# Semantic Web solutions

The Semantic Web takes the solution further. It involves publishing in languages specifically designed for data: Resource Description Framework (RDF). Web Ontology Language (OWL), and Extensible Markup Language (XML). HTML describes documents and the links between them. RDF, OWL, and XML, by contrast, can describe arbitrary things such as people, meetings, or airplane parts. Tim Berners-Lee calls the resulting network of Linked Data the Giant Global Graph, in contrast to the HTML-based World Wide Web.

These technologies are combined to provide descriptions that supplement or replace the content of Web documents. Thus, content may manifest itself as descriptive data stored in Web-accessible databases, or as markup within documents (particularly, in Extensible HTML (XHTML) interspersed with XML, or, more often, purely in XML, with layout or rendering cues stored separately). The machine-readable descriptions enable content managers to add meaning to the content, i.e., to describe the structure of the knowledge we have about that content. In this way, a machine can process knowledge itself, instead of text, using processes similar to human deductive reasoning and inference, thereby obtaining more meaningful results and helping computers to perform automated information gathering and research.

# Relationship to object-oriented programming

Several authors highlight the similarities that the Semantic Web shares with object-oriented programming (OOP). Both the semantic web and object-oriented programming have classes with attributes and the concept of instances or objects. Linked Data uses Dereferenceable Uniform Resource Identifiers like the common programming concept of pointers or” object identifiers” in OOP. Dereferenceable URIs can thus be used to access” data by reference”. The Unified Modeling Language is designed to communicate about object-oriented systems, and can thus be used for both object-oriented programming and semantic web development.

# Challenges

Some of the challenges for the Semantic Web include vastness, vagueness, un- certainty, inconsistency, and deceit. Automated reasoning systems will have to deal with all of these issues to deliver on the promise of the Semantic Web.

## Vastness

The World Wide Web contains at least 48 billion pages as of this writing (August 2, 2009). The SNOMED CT medical terminology ontology contains 370,000 class names, and existing technology has not yet been able to eliminate all semantically duplicated terms. Any automated reasoning system will have to deal with truly huge inputs.

## Vagueness

These are imprecise concepts like ”young” or ”tall”. This arises from the vagueness of user queries, of concepts represented by content providers, of matching query terms to provider terms, and of trying to combine different knowledge bases with overlapping but subtly different concepts. Fuzzy logic is the most common technique for dealing with vagueness.

## Uncertainty

These are precise concepts with uncertain values. For example, a patient might present a set of symptoms that correspond to several different distinct diagnoses each with a different probability. Probabilistic reasoning techniques are generally employed to address uncertainty.

## Inconsistency

These are logical contradictions that will inevitably arise during the development of large ontologies, and when ontologies from separate sources are combined. Deductive reasoning fails catastrophically when faced with inconsistency, because ”anything follows from a contradiction”. Defeasible reasoning and paraconsistent reasoning are two techniques that can be employed to deal with inconsistency.

## Deceit

This is when the producer of the information is intentionally misleading the consumer of the information. Cryptography techniques are currently utilized to alleviate this threat.