

The Semantic Web: Key Ideas

Grigoris Antoniou, Vassilis Christophides, Dimitris Plexousakis
Institute of Computer Science, FORTH, Greece

INTRODUCTION

World Wide Web: A Critical View

The World Wide Web (Berners-Lee, Cailliau & Groff, 1992; Berners-Lee, 1999) has changed the way people communicate with each other and the way business is conducted. It lies at the heart of a revolution which is currently transforming the developed world towards a knowledge economy (Neef, 1997), and more broadly speaking, to a knowledge society.

Most of today's Web content is suitable for *human consumption*. Even Web content that is generated automatically from data bases is usually presented without the original structural information found in data bases. Typical uses of the Web today involve humans seeking and consuming information, searching and getting in touch with other humans, reviewing the catalogs of online stores and ordering products by filling out forms.

These activities are not particularly well supported by software tools. Apart from the existence of links which establish connections between documents, the main valuable, indeed indispensable, kind of tools are *search engines*. Keyword-based search engines, such as AltaVista (www.altavista.com), Yahoo (www.yahoo.com) and Google (www.google.com; Page and Brin, 1998), are the main tool for using today's Web. It is clear that the Web would not have been the huge success it was, were it not for search engines. However there are serious problems associated with their use; the most important ones are listed in Table 1.

Table 1: A List of Problems Associated with Keyword-based Search

High recall, low precision

Too many, mostly irrelevant pages are retrieved

Low or no recall

Key relevant pages are not retrieved

Sensitivity to chosen vocabulary

Slight changes in vocabulary may cause significant changes in result

Results are single Web pages

Information may be spread across various pages

Human involvement is necessary

To interpret retrieved pages, and to combine information

BACKGROUND

The Semantic Web Vision

The Semantic Web (Berners-Lee, Hendler & Lassila, 2001; Davis, Fensel and van Harmelen, 2002; Fensel et al., 2002; Antoniou & van Harmelen, 2004) proposes to overcome the difficulties listed above by making Web content *machine-processable*. The key point is that the *semantics* (meaning) of Web content is explicitly represented and processed. This aim will be achieved by the combination of the following technologies:

- *Explicit Meta-Data*: Web content will carry its meaning “on its sleeve” through appropriate semantic markup.
- *Ontologies*: They will describe semantic relationships between terms, and will serve as the foundation for establishing shared understanding between applications
- *Logical Reasoning*: Automated reasoning-enabled tools will make use of the information provided by meta-data and ontologies.

More on ontologies and reasoning are found in the following sections. As a simple example, suppose that you are searching for photos of an orange ape in an annotated online collection of digital photos. Suppose that picture 1 is annotated as “playing orangutan”. Then this picture can be retrieved on the Semantic Web although its annotation does not contain the words “orange” or “ape” (so a keyword-based search would fail). This can be achieved through interplay of (i) the annotation, (ii) information contained in an ontology about animals which states that orangutans are apes and orange, and (iii) reasoning which combines the above information to conclude that the picture is relevant to the user’s query.

The Semantic Web vision was created by Tim Berners-Lee, the very person who created the WWW. The Semantic Web activities are coordinated by the World Wide Web Consortium (<http://www.w3.org/2001/sw/>). Table 2 collects a few critical issues of Web technologies addressed by the Semantic Web initiative.

Table 2: Critical Issues of Web Technologies (Khrosow-Pour 2005) addressed by the Semantic Web

Cyberloafing

Surfing the Internet, wasting time and accessing inappropriate materials

Flooding of the Web with content

Including information that is not helpful

Inadequate search facilities on the WWW

Lack of a high level query language search engines for locating, filtering, and presenting information

Maintaining and integrity of data

Maintaining up-to-date and accurate information on the site for viewers to use

FUTURE TRENDS

Ontologies

An ontology is a formalization of a shared conceptualization of a particular domain. It supports interoperability between applications at the semantic level (the meaning of information), but also allows people to talk about objects of common interest. Typically it describes (a) the objects of the domain; (b) relationships that hold between them.

Objects of the same kind are organized in so-called *classes*: collections of objects sharing certain characteristics. For example, in a university domain, classes may be professors, students, administrative and technical staff, courses, lecture theaters etc. Individual objects (called *resources* in Web terminology) can be declared to be *instances* of a certain class.

Once we have defined classes we can also define relationships between them. One particular kind of relationships are *class hierarchies*. A class A is subclass of a class B (and B a superclass of A) if every instance of A is also an instance of B. For example, the class of professors is a subclass of the class of all university employees. Figure 1 shows a sample class hierarchy for the university domain.

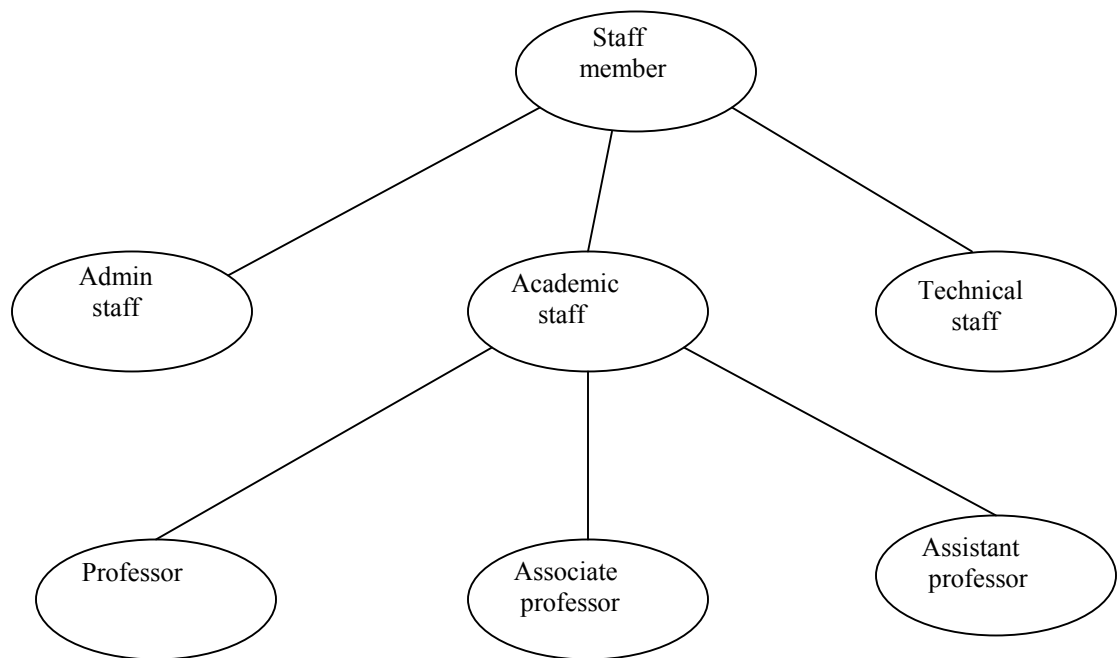


Figure 1: A Class hierarchy in the university domain

The subclass relationship is a general-purpose relationship between classes. Further relationships, called *properties*, can be defined by the user. In the university domain, such properties might be that a course is taught by a

particular professor, that a student takes a course, and that a professor is Head of a department.

The interplay of classes and properties opens interesting modeling possibilities. For one, it is possible to define domain and range restrictions. For example, one can specify that a course can be only taught by an academic staff member (*range restriction*) and that only a course, and not, say, a lecture theater, can be taught (*domain restriction*).

Moreover, the valuable concept of *inheritance* can be utilized. Suppose that person X is declared to be an associate professor. Then X is allowed to teach a course because he inherits this possibility from its superclass academic staff member. This way we can avoid adding superfluous information to the ontology (X is an academic staff member; X is a staff member).

Information such as the above can be expressed in an ontology language. Ontology languages can be rather informal, e.g. thesauri, or formal when they rely on some kind of logical formalism.

Ontologies have nowadays become valuable tools in various domains, for example in biology [Bodenreider, Mitchell & McCray, 2003], medicine [UMLS], and culture [CIDOC/CRM].

Reasoning

When an ontology language is formal, it is possible to reason about the knowledge expressed in a particular ontology. For example, we may reason about:

- Class membership: If x is an instance of class C, and C is a subclass of D, then we can infer that x is an instance of D.
- Equivalence of classes: If class A is equivalent to class B, and B is equivalent to class C, then we can infer that A is equivalent with C.
- Consistency: If we have declared that classes C and D are disjoint, and x is an instance of both C and D, then there is an error.
- Classification: If we have declared that certain property-value pairs are sufficient conditions for membership in class A, then if an individual x satisfies such conditions, we can conclude that x must be an instance of A.

Derivation such as the preceding can be made mechanically instead of being made by hand. Such reasoning support is important because it allows one to

- check the consistency of an ontology and the knowledge
- check for unintended between classes
- automatically classify instances of classes

Automated reasoning support allows one to check many more classes than could be checked manually. Checks like the preceding ones are valuable for designing large ontologies, where multiple authors are involved, and for integrating and sharing ontologies from various sources.

Relevance to Sample Application Fields

Tables 3 and 4 illustrate the relevance of the Semantic Web vision to the application areas of Knowledge Management and B2C E-Commerce (Fensel, 2001). Other areas that would greatly benefit from the realization of the Semantic Web vision include B2B E-Commerce, E-Learning (Nilsson, Palmer & Naeve, 2002), E-Government (Klischewski, 2003), Web services (McIlraith, Son & Zeng, 2001) and Peer-to-Peer systems (Nejdl et al., 2002).

Table 3: The Semantic Web and Knowledge Management

- Knowledge will be organised in conceptual spaces, that is, according to concepts and their interrelations, not according to keywords or other ad hoc means.
- Automated tools will support maintenance by checking for inconsistencies and extracting new knowledge, hidden in the stored knowledge.
- Keyword-based search will be replaced by query answering based on what needs to be retrieved (semantic query answering).
- Definition of views on certain parts of information (even parts of documents) will be possible. This way, certain pieces of knowledge will be hidden from certain groups of persons, while they will be accessible to others.

Table 4: The Semantic Web and B2C E-Commerce

- Pricing and product information will be extracted correctly, delivery and privacy policies will be interpreted and compared to the user requirements.
- Additional information about the reputation of online shops will be retrieved from other sources, for example, independent rating agencies or consumer bodies.
- The low-level programming of wrappers, necessary for today's shopbots, will become obsolete.
- More sophisticated shopping agents will be able to conduct automated negotiations, on the buyer's behalf, with shop agents.

C O N C L U S I O N

The Semantic Web is the next step in the development of the WWW. It proposes to dramatically improve the Web's functionality and usability by making the meaning of Web content machine processable.

At present, most of the effort has been placed on the design of formal languages for describing Web resources and for writing Web ontologies; the most important languages are the language of RDF (Lassila & Swick, 1999) which allows one to state simple facts, and the ontology languages RDF Schema (Brickley & Guha, 2003) and OWL (Dean & Schreiber, 2003). The success of the Semantic Web will critically rely on two factors:

1. The development of professional tools for annotating pages and linking them to existing ontologies, for developing ontologies, and for reasoning with them. RDF Suite (Karvounarakis et al., 2002) is a good collection of tools supporting RDF and RDF Schema.
2. The uptake by users. Initial successes can be expected either in environments with central authority (e.g. in the knowledge management of large organizations), or in the emergence of virtual communities.

R E F E R E N C E S

- Antoniou, G. and van Harmelen, F. (2004). *A Semantic Web Primer*. Cambridge: The MIT Press.
- Berners-Lee, T.J., Calilliau and Groff, J.F. (1992). The World Wide Web. *Computer Networks & ISDN Systems*, 25(4), 454-459.
- Berners-Lee, T. (1999). *Weaving the Web*. Harper.
- Berners-Lee, T., Hendler, J. and Lassila, O. (2001). The Semantic Web. *Scientific American*, 284(5), 34-43.
- Bodenreider, O., Mitchell, J.A. and McCray, A.T. (2003). *Biomedical Ontologies*. In: *Proceedings 2003 Pacific Symposium on Biocomputing*. World Scientific 2003, 562-564.
- Brickley, D. and Guha, R.V. (2003). *RDF Vocabulary Description Language 1.0: RDF Schema*. <http://www.w3.org/TR/rdf-schema/>
- CDOC/CRM. *CIDOC Conceptual Reference Model*. <http://cidoc.ics.forth.gr>
- Davis, J., Fensel, D. and van Harmelen, F. (2002). *Towards the Semantic Web: Ontology-Driven Knowledge Management*. Wiley.
- Dean, M. and Schreiber, G. (2003). *OWL Web Ontology Language Reference*. <http://www.w3.org/TR/owl-ref/>
- Fensel, D. (2001). *Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce*. Berlin: Springer-Verlag.
- Fensel, D., Hendler, J., Lieberman, H. and Wahlster, W. (2003). *Spinning the Semantic Web*. Cambridge: The MIT Press.
- Karvounarakis, G., Alexaki, S., Christophides, V., Plexousakis, D. and Scholl, M. (2002). RQL: A Declarative Query Language for RDF. *Proceedings of the 11th International World Wide Web Conferenc (WWW'02)*.

Khosrow-Pour, M. (2005). An Overview of Web-Enabled Technologies Assessment and Management: Critical Issues, *Encyclopedia of Information Science and Technology*, Mehdi Khosrow-Pour (ed.), Hershey, PA: Idea Group Publishing.

Klischewski, R. (2003). Semantic Web for e-Government, In: Traunmüller, R. (ed.): *Proceedings of EGOV'2003*. Berlin: Springer, LNCS 2739, 288-295.

Lassila, O. and Swick, R.R. (1999). *Resource Description Framework (RDF) Model and Syntax Specification*. <http://www.w3.org/TR/REC-rdf-syntax/>

McIlraith, S.A., Son, T.C. and Zeng, H. (2001). Semantic Web Services. *IEEE Intelligent Systems* 16(2): 46-53.

Neef, D. (1997). *The Knowledge Economy*. Butterworth-Heinemann.

Nejdl, W., Wolf, B., Qu, C., Decker, S., Sintek, M., Naeve, A., Nilsson, M., Palmér, M. and Risch, T. (2002). EDUTELLA: a P2P networking infrastructure based on RDF. *Proceedings of the 11th International World Wide Web Conference (WWW'2002)*, 604-615.

Nilsson, M., Palmér, M. and Naeve, A. (2002). Semantic Web Meta-data for e-Learning - Some Architectural Guidelines. *Proceedings of the 11th World Wide Web Conference (WWW2002)*, Hawaii, USA.

Page L. and Brin S. (1998). The Anatomy of a Search Engine. In *Proc. 7th International WWW Conference (WWW'98)*. Brisbane, Australia, April 14-18, 1998.

UMLS. *Unified Medical Language System*.
<http://www.nlm.nih.gov/research/umls/umlsmain.html>

Terms and Definitions

Class: A collection of objects sharing certain characteristics (such as producer, customer, product, company etc.) Individual objects are called instances of the class. The classes of an ontology describe the important concepts of a particular domain being modeled.

Class hierarchy: Classes are usually organized in a conceptual space along a generalization/specialization axis. A class A is more general (superclass) than a class B when each instance of B is also instance of A.

Ontology: A consensual and formal description of shared concepts in a domain. Typically it organizes the objects of a domain in classes, and includes a hierarchy of classes (e.g. printers are Hardware devices). Ontologies are used to aid human communication and shared understanding, but also communication among software applications.

OWL: The current W3C standard for defining Web ontologies.

Property: Properties are used to establish links between classes. For example, books are published by publishers. The property “publishes” relates a publisher with a particular book.

Property hierarchy: As with classes, properties can be organized in generalization/specialization taxonomies. For example, the property “is CEO of” relating persons to companies is a more specific property than (subproperty of) “works for”.

RDF: The basic language of the Semantic Web. It is used for describing Web resources.

RDF Schema: A primitive ontology language, integrated with RDF. Its basic operations are the definition of classes and properties, their organization in hierarchies, and domain and range restrictions on properties.

XML: An application-independent meta-language for defining markup languages. It serves as the basis for syntactic interoperability among information systems.