

Semantic Web

The **Semantic Web** is a project that intends to create a universal medium for information exchange by giving meaning (semantics), in a manner understandable by machines, to the content of documents on the Web. Currently under the direction of the Web's creator, Tim Berners-Lee of the World Wide Web Consortium, the Semantic Web extends the ability of the World Wide Web through the use of standards, markup languages and related processing tools.

Relationship to the World Wide Web

Currently, the World Wide Web is based primarily on documents written in HTML, a language that is useful for describing, with an emphasis on visual presentation, a body of structured text interspersed with multimedia objects such as images and interactive forms. HTML has limited ability to classify the blocks of text on a page, apart from the roles they play in a typical document's organization and in the desired visual layout.

For example, with HTML and a tool to render it (perhaps Web browser software, perhaps another user agent), one can create and present a page that lists items for sale. The HTML of this catalog page can make simple, document-level assertions such as "this document's title is 'Widget Superstore'". But there is no capability within the HTML itself to unambiguously assert that, say, item number X586172 is an Acme Gizmo with a retail price of €199, or that it is a consumer product. Rather, HTML can only say that the span of text "X586172" is something that should be positioned near "Acme Gizmo" and "€199", etc. There is no way to say "this is a catalog" or even to establish that "Acme Gizmo" is a kind of title or that "€199" is a price. There is also no way to express that these pieces of information are bound together in describing a discrete item, distinct from other items perhaps listed on the page.

The Semantic Web addresses this shortcoming, using the descriptive technologies RDF and OWL, and the data-centric, customizable markup language XML. These technologies are combined in order to provide descriptions that supplement or replace the content of Web documents. Thus, content may manifest as descriptive data stored in Web-accessible databases, or as markup within documents (particularly, in XHTML interspersed with XML, or, more often, purely in XML, with layout/rendering cues stored separately). The machine-readable descriptions allow content managers to add meaning to the content, thereby facilitating automated information gathering and research by computers.

Components of the Semantic Web

The Semantic Web is comprised of the standards and tools of XML, XML Schema, RDF, RDF Schema and OWL. The OWL Web Ontology Language Overview describes the function and relationship of each of these components of the Semantic Web:

- XML provides a surface syntax for structured documents, but imposes no semantic constraints on the meaning of these documents.
- XML Schema is a language for restricting the structure of XML documents.
- RDF is a datamodel for objects ("resources") and relations between them, provides a simple semantics for this datamodel, and these datamodels can be represented in an XML syntax.
- RDF Schema is a vocabulary for describing properties and classes of RDF resources, with a semantics for generalization-hierarchies of such properties and classes.
- OWL adds more vocabulary for describing properties and classes: among others, relations between classes (e.g. disjointness), cardinality (e.g. "exactly one"), equality, richer typing of properties, characteristics of properties (e.g. symmetry), and enumerated classes.

The intent is to enhance the usability and usefulness of the Web and its interconnected resources through:

- documents "marked up" with semantic information (an extension of the HTML <meta> tags used in today's Web pages to supply information for Web search engines using web crawlers). This could be machine-readable information about the human-readable content of the document (such as the creator, title, description, etc., of the document) or it could be purely metadata representing a set of facts (such as resources and services elsewhere in the site). (Note that *anything* that can be identified with a *Uniform Resource Identifier (URI)* can be described, so the semantic web can reason about people, places, ideas, cats etc.)
- common metadata vocabularies (ontologies) and maps between vocabularies that allow document creators to know how to mark up their documents so that agents can use the information in the supplied metadata (so that **Author** in the sense of 'the Author of the page' won't be confused with **Author** in the sense of a book that is the subject of a book review).
- automated agents to perform tasks for users of the Semantic Web using this metadata
- web-based services (often with agents of their own) to supply information specifically to agents (for example, a Trust service that an agent could ask if some online store has a history of poor service or spamming).

The primary facilitators of this technology are URIs (which identify resources) along with XML and namespaces. These, together with a bit of logic, form RDF, which can be used to say anything about anything. As well as RDF, many other technologies such as Topic Maps and pre-web artificial intelligence technologies are likely to contribute to the Semantic Web. A popular application of the Semantic Web is Friend of a Friend (or Foaf), which describes people and other agents in terms of RDF.