The Semantic Web

The article **"The Semantic Web"** by Berners-Lee, Hendler, and Lassila (2001)[1] introduces the vision of the Semantic Web as a system that allows computers to infer meaning from the relationships between resources in the Web. Currently most of the Web's content is designed for humans to read, and computers can identify which part of the content is a header, or a graphic or link. However, they cannot reliably understand the meaning, or 'semantics', of the text.

The Semantic Web, an extension of the existing Web, is a way of structuring content to enable software agents to carry out tasks that require a level of logical understanding. For instance, it is easy for us to search online for the name of a doctor's clinic or surgery. We can then click through to scan the content of the page for details such as opening times, how to book an appointment, etc. With the Semantic Web, this information can be returned to us by agents. To enable computers to understand content, we must give information well-defined meaning.

Knowledge Representation

Knowledge Representation systems have been used in artificial intelligence for decades and are structured collections of information and sets of inference rules that can be used to conduct automated reasoning. The challenge of the Semantic Web is to provide a language that allows rules from any existing knowledge-representation system to be exported onto the Web.

Much of the information we need to express is of a simple nature, such as 'an obstetrician is a kind of doctor'. To do this, we use eXtensible Markup Language (XML) and the Resource Description Framework (RDF).

XML lets people create their own tags—hidden labels such as <address> that annotate Web pages or sections of text on a page, much like tags in HTML. This adds structure to the page, but does not tell a computer what the meaning of

[1]Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web.*Scientific American*, *284*(5), 28-37.

that structure is. For this, we use RDF. In RDF, a document states that particular things (people, places) have properties (such as "is a sister of," "is the capital of") with certain values (another person, a country). These are known as 'triples' and vast majority of data processed by machines can be expressed in this way.

Subject and value are each identified by a Universal Resource Identifier (URI), of which URLs are the most common form. Each URI has a unique definition, so we can distinguish between the actor Anne Hathaway and Shakespeare's wife, Anne Hathaway. The relationships are also identified by URIs, which enables anyone to define a new concept by creating a URI for it on the Web.

Ontologies

A problem arises when two people have used different identifiers for the same concept. How do we know these are in fact referring to the same thing? In Web Science, an ontology is a file that describes the relationship amongst terms. Typically, a Semantic Web ontology has a 'taxonomy' and a set of inference rules.

The taxonomy defines classes of objects and relations among them. If we assign properties to classes and allow subclasses to inherit those properties, we can show how they are related. If city code is of type city, then we can look up the website associated with the city code even if there is no such link, because we can find the website associated with the city.

Having established the terms and relationships, inference rules allow us to use those relationships to arrive at logical conclusions, such as, "If a house is on a street, and that street is in a specific town, then the house must be in the town" even if we do not have the town in the address.

Ontologies also mean that it is easier to develop programmes that pull together information from several different sources.

Agents

The real power of the Semantic Web will be realized when people create many programs that collect Web content from diverse sources, process the information and exchange the results with other programmes. The Semantic Web enables even agents that were not designed to work together to transfer data seamlessly.

Naturally, an important part of the Semantic Web is that we can trust that it is returning the correct information to us if we are not checking the content of a document ourselves. This is provided by agents exchanging "proofs" written in the Semantic Web's unifying language (the language that expresses logical inferences made using rules and information such as those specified by ontologies).

Another vital feature will be digital signatures, which are encrypted blocks of data that computers and agents can use to verify that the attached information has been provided by a specific trusted source.

The Semantic Web also enables usage of the Internet of Things (IoT). URIs can point to anything, including physical entities. Using RDF to describe things such as cell phones and TVs enables us to control many devices with one simple configuration.

Evolution of knowledge

By enabling anyone express new concepts that they invent with URIs, and link them logically, the Semantic Web enables new concepts to be progressively linked into a universal Web. This structure will open up the knowledge and workings of humankind to meaningful analysis by software agents, providing a new class of tools by which we can live, work and learn together.