Healthcare Information Systems and the Semantic Web

David Parry
Auckland University of Technology, New Zealand

INTRODUCTION

The World Wide Web (WWW) is a critical source of information for health care. Because of this, systems for allowing increased efficiency and effectiveness of information retrieval and discovery are critical. Increased intelligence in Web pages will allow information sharing and discovery to become vastly more efficient .The semantic Web is an umbrella term for a series of standards and technologies that will support this development.

BACKGROUND

The Web and Healthcare Information Systems

The early development of the Internet, from ARPA developments and the Internet protocol in the 1960s and 1970s to increasing use of e-mail and text-based systems, led to the development of HTTP and HTML in the 1990s (Barry et al., 1997). The Web has become a critical source of information and communication among people in all domains. The world of medicine on the Web, including such concepts as eHealth, have become areas of great academic interest (Pagliari et al., 2005). Thus, the Web and health care are increasingly co-dependent.

Finding the required information in a search of HTML documents is difficult, as HTML is really a formatting language for humans rather than for indexing. Of course, Web-based systems such as search engines or digital libraries are available, but these rely on indexing either as a set of keywords drawn from a limited vocabulary or from sometimes unreliable parsing of the document. Knowing what information a Web page contains, whether it is a medical record, a clinical result, or an academic paper, is still difficult.

The Semantic Web

Tim Berners-Lee has coined the term "Semantic Web" (Berners-Lee, Hendler, & Lassila, 2001). This term has been enthusiastically adopted and generally includes any work intended to allow the meaning of data contained in Web resources to be made accessible to software agents or humans.

In essence, a semantic Web document is a collection of three parts:

- The document that contains the information, marked up in a suitable way
- A document that contains the rules for interpreting the markup (Namespace)
- A document that links the information with the namespace (i.e., which set of interpretation rules is being used).

There has been a great deal of work under the auspices of W3C to develop standards for the semantic Web, in particular representations using XML. Two of the major efforts are the Web ontology language (OWL) (Smith, Welty, & McGuinness, 2004) and the resource description framework (RDF). In terms of information retrieval, the semantic Web will allow a user to express what they mean to find and find objects that satisfy that request without regard to the language or syntax of the request. The semantic Web implies ways of representing the meaning of documents and constructing queries to discover that meaning. If possible, such a process will be as automatic as possible and incorporate the advances in computing power available to users on the desktop and at the server.

An example may make this system clearer. The example is taken from the W3C example page (http://www.w3.org/2000/10/rdf-tests/)

The original RDF document is:

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:a="http://description.org/schema/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:oiled="http://img.cs.man.ac.uk/oil/oiled#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:daml="http://www.daml.org/2001/03/daml+oil#"
  xml:base="file:/C:/downloads/RDF-Model-Syntax_1.0/
ms_4.1_1.rdf">
 <rdf:Statement>
  <rdf:subject rdf:resource="http://www.w3.org/Home/Las-
sila"/>
 <rdf:predicate rdf:resource="http://description.org/schema/
Creator"/>
  <a:attributedTo>Ralph Swick</a:attributedTo>
  <rdf:object>Ora Lassila</rdf:object>
 </rdf:Statement>
</rdf:RDF>
```

This can be displayed as shown in Figure 1.

This can be expressed in words as "Ralf Swick and Ora Lassila have the attributes of creators of RDF."

The flexibility of this approach can be seen by the fact that the definitions of what RDF is and what "creator" means can be located in other files. Thus, agreed meanings can be used across multiple data files.

Often it may be useful to think of the meanings of items in documents as related to an ontology and that there is a standard for the expression of ontologies on the WWW. This standard—Ontology Web Language (OWL) (Smith, Welty, & McGuinness, 2004)—allows ontology relations to be coded in a machine and human readable form within files that can be located on the WWW.

An example of an OWL statement is shown in Figure 2 (the example is taken from Smith, Welty, & McGuinness, 2004.

This example shows the definition of an equivalent class—items that satisfy the conditions given in the equivalent class construction are regarded as equivalent. In this particular example, things from Texas are assigned to be "Texas things." This project is still at a fairly early stage but obviously represents a potentially beneficial approach to standardization that may be useful for the deployment of the fuzzy ontology described later.

What both these approaches do is allow the page author to identify to agents or humans the intended meaning of various parts of the document. The RDF form does not need the existence of an underlying ontology, but in many cases, an ontology does exist, which makes consistent namespace creation easier. Of course, systems such as HL7 already have what are effectively namespaces, but the semantic Web approach means that the author provides the data, the possible values of the data, and the potential meaning of the data in a format that allows shared meanings. This

Figure 1. XML document construction