

Monday, January 13, 2003

I've been studying [RDF](#) and the [Semantic Web](#) a lot recently. I've taken stabs at learning it several times over the past two years, and each time it makes a little more sense to me.

Let me give an introduction to the concepts behind RDF, as I've come to understand them:

RDF is basically a framework for describing resources. A resource is anything that can be identified by a URI, such as a web page. For instance, you might say "This web page is authored by Jason Johnston". Subject, verb, object. This is what they mean by "Semantic Web": everything on the Web could potentially be described and tied together by similar statements.

Now, RDF itself is only a way of saying that relationships exist between resources. The *types* of relationships themselves are described using languages defined elsewhere, such as [Dublin Core](#) (a standard metadata format) or any language you wish to invent. So in theory as long as a language exists to define the meaning of a particular relationship, then you can use RDF to say that a resource has that relationship with something. RDF is just a way of putting the sentence together.

As an example, using the RDF framework you could say something like:

“The resource <http://www.w3.org/2001/sw> has a property 'Title' from the namespace http://purl.org/metadata/dublin_core# with the value 'W3C Semantic Web'.”

The web page resource is tied to the value "W3C Semantic Web" by the "Title" property. We know what "Title" means because it is defined by the Dublin Core metadata language (identified by its unique namespace). Because it has a URI, a property therefore is also itself a resource. So we have two types of objects in RDF: resources (such as the web page and the namespaced "Title" property) and literals (such as the actual title string). Here are the parts of the statement:

Weblog Entry for Monday, January 13, 2003

- Subject: the resource <http://www.w3.org/2001/sw>
- Property: the resource http://purl.org/metadata/dublin_core#Title
- Value: the literal "W3C Semantic Web"

And here is how it might be encoded in [RDF XML Serialization Syntax](#):

```
<rdf:Description about="http://www.w3.org/2001/sw"
xmlns:dc="http://purl.org/metadata/dublin_core#"> <dc:Title>W3C
Semantic Web</dc:Title> </rdf:Description>
```

The "Value" in the above statement doesn't have to be a literal, it can also be a resource with its own URI. In this way we can describe relationships *between* resources, such as "This web page has an author represented by his bio at that other web page." And thus the entire Web is described and connected semantically by simple sentences.

What's the point, you ask? Currently one of the weaknesses of the Web is that resources on it are largely unconnected in any meaningful way. Sure, you have HTML links which do connect documents to some extent, but when you follow a link how much do you *really* know about how the target resource is related to the resource you're currently viewing? The text within the link might provide some clue, but that is arbitrary and unreliable. What's more, interpreting link text requires a human, whereas RDF can be machine-interpreted.

By offering a unified method for describing explicit, well-defined relationships between resources RDF could potentially turn the Web into one world-wide distributed database. Imagine going to your favorite Semantic Web Search Engine and easily finding a list of biographies of authors of online novels published the same month as a certain product in an online catalog received its patent. This will be a reality because all that metadata is linked together in a standardized way.

Posted by Jason at [14:56:28](#) - [[0 Comments](#)]