NOTE-rdf-simple-intro-971113.html

Introduction to RDF Metadata

W3C NOTE 1997-11-13

This document: http://www.w3.org/TR/NOTE-rdf-simple-intro-971113.html

Author: Ora Lassila, ora.lassila@research.nokia.com, Nokia Research Center

---

Status of this Document

This document is a NOTE made available by the W3 Consortium for discussion only. This indicates no endorsement of its content, nor that the Consortium has, is, or will be allocating any resources to the issues addressed by the NOTE.

Thie document provides a brief introduction to the Resource Description Framework (RDF) and the concept of metadata, and is intended as "prerequisite reading" for those trying to understand the RDF specification. An earlier version was written for Nokia Research Center's internal journal *Advance*.

Current Issues

One of the major issues of the World Wide Web as it exists today is that it is really hard to automate any tasks which one has to perform on the web. So far, the web is mainly built as a forum for human interaction; because most web documents are written for human consumption, the only available form of searching on the web (for example) is to simply match words or sentences contained in documents. Anyone who has used a web search service like AltaVista or HotBot knows that typing in a few keywords and receiving a couple of thousand "hits" is not necessarily very useful. A lot of manual "weeding" of information has to happen after that; it may also happen that the keywords for which you are searching are not prominent in the relevant document itself.

A possible solution for the search problem - and for the general issue of letting automated "agents" roam the web performing useful tasks - is to provide a mechanism which allows a more precise description of things on the web. This, in turn, could elevate the status of the web from *machine-readable* to something we might call *machine-understandable*.

Metadata is "data about data" or specifically in our current context "data describing web resources." The distinction between "data" and "metadata" is not an absolute one; it is a distinction created primarily by a particular application ("one application's metadata is another application's data").

Standardization Efforts at W3C

One could say that the history of metadata at W3C begins with PICS - or Platform for Internet Content Selection. PICS is a mechanism for communicating ratings of web pages from a server to clients; these ratings, or *rating labels*, contain information about the content of web pages: for example, whether a particular page contains a peer-reviewed research article, or was authored by an accredited researcher, or contains sex, nudity, violence, foul language etc. Instead of being a fixed set of criteria, PICS introduced a general mechanism for creating *rating systems*. Different organizations could rate content based on their own objectives and values, and users - for example, parents worried about their children's web usage - could set their browser to filter out any web pages not matching their own criteria. Development of PICS was motivated by the anticipation of restrictions on the Internet such as some recent US legislation (the Communications Decency Act and its subsequent overruling by the Federal Supreme Court).

PICS is a restricted metadata framework. It allows certain things to be expressed very precisely about web pages; in particular, PICS is useful when all the possible data values can be known in advance. The development of RDF as a general metadata framework - and in a way as a general *knowledge representation mechanism* for the web - was heavily inspired by PICS.

RDF - the Resource Description Framework, as our proposed mechanism is called - is a foundation for processing metadata; it provides interoperability between applications that exchange machine-understandable information on the Web. RDF emphasizes facilities to enable automated processing of Web resources. RDF metadata can be used in a variety of application areas; for example: in *resource discovery* to provide better search engine capabilities; in *cataloging* for describing the content and content relationships available at a particular Web site, page, or digital library; by *intelligent software agents* to facilitate knowledge sharing and exchange; in *content rating*; in describing *collections* of pages that represent a single logical "document"; for describing *intellectual property rights* of Web pages, and in many others. RDF with *digital signatures* will be key to building the "Web of Trust" for electronic commerce, collaboration, and other applications.

RDF encourages the view of "metadata being data" by using XML (the eXtensible Markup Language) as its encoding syntax. The resources being described by RDF are, in general, anything that can be named via a URI (Uniform Resource Identifier). The broad goal of RDF is to define a mechanism for describing resources that makes no assumptions about a particular application domain, nor defines the semantics of any application domain. The definition of the mechanism should be domain neutral, yet the mechanism should be suitable for describing information about any domain.

The recently published document about RDF introduces a model for representing metadata and one possible syntax for expressing and transporting this metadata in a manner that maximizes the interoperability of independently developed web servers and clients. This document is to be followed by others addressing issues such as how to define *schemata* (classes) for metadata, how to write queries, etc.

So What Is RDF Like, Really?

At the core, RDF data consists of *nodes* and attached *attribute/value pairs*. Nodes can be any web resources (pages, servers, basically anything for which you can give a URI), even other instances of metadata. Attributes are named properties of the nodes, and their values are either *atomic* (text strings, numbers, etc.) or other resources or metadata instances. In short, this mechanism allows us to build

labeled directed graphs.

The essence of RDF is the model of nodes, attributes, and their values. In order to store instances of this model into files or to communicate these instances from one agent to another, we need a graph serialization syntax. The particular language we use is XML (XML being W3C's work-in-progress to define a richer Web syntax for a variety of applications). RDF and XML are complementary; there will be alternate ways to represent the same RDF data model, some more suitable for direct human authoring.

RDF in itself does not contain any predefined vocabularies for authoring metadata. We do, however, expect that standard vocabularies will emerge, after all this is a core requirement for large-scale interoperability. Some of the vocabularies in the foreseeable future are a PICS-like rating architecture, a digital library vocabulary (currently referred to as "Dublin Core"), and a vocabulary for expressing digital signatures. Anyone can design a new vocabulary, the only requirement for using it is that a designating URI is included in the metadata instances using this vocabulary. This use of URIs to name vocabularies is an important design feature of RDF: many previous metadata standardization efforts in other areas have foundered on the issue of establishing a central attribute registry. RDF permits a central registry but does not require one.

Future of Metadata on the Web

The RDF *working group* - the W3C vehicle for crafting new standards - includes representatives from key companies and organizations: Netscape, Microsoft, IBM, Nokia, OCLC, etc. The interest from the large web browser vendors gives us hope that large scale deployment of tools which understand about RDF will take place; this in turn should lead to the widespread adoption of RDF on the web.

Once the web has been sufficiently "populated" with rich metadata, what can we expect? First, searching on the web will become easier as search engines have more information available, and thus searching can be more focused. Doors will also be opened for automated software agents to roam the web, looking for information for us or transacting business on our behalf. The web of today, the vast unstructured mass of information, may in the future be transformed into something more manageable - and thus something far more useful.

More Information

- A press release from W3C.
- The RDF home page.
- The current RDF model and syntax proposal.

Ora Lassila <ora.lassila@research.nokia.com>