Resource Description Framework

From Wikipedia, the free encyclopedia

Resource Description Framework

Current Status	Published, W3C Recommendation
Editors	Frank Manola, Eric Miller
Base Standards	XML, URI
Related Standards	RDFS, OWL, RIF, RDFa
Domain	Semantic Web
Abbreviation	RDF
Website	RDF Primer

The **Resource Description Framework** (**RDF**) is a family of World Wide Web Consortium (W3C) specifications [1] originally designed as a metadata data model. It has come to be used as a general method for conceptual description or modeling of information that is implemented in web resources, using a variety of syntax formats.

Contents

[hide]

[edit] Overview

The RDF data model[2] is similar to classic conceptual modeling approaches such as entity-relationship or class diagrams, as it is based upon the idea of making statements about resources (in particular Web resources) in the form of subject-predicate-object expressions. These expressions are known as *triples* in RDF terminology. The subject denotes the resource, and the predicate denotes traits or aspects of the resource and expresses a relationship between the subject and the object. For example, one way to represent the notion "The sky has the color blue" in RDF is as the triple: a subject denoting "the sky", a predicate denoting "has the color", and an object denoting "blue". RDF is an abstract model with several serialization formats (i.e., file formats), and so the particular way in which a resource or triple is encoded varies from format to format.[3]

This mechanism for describing resources is a major component in what is proposed by the W3C's Semantic Web activity: an evolutionary stage of the World Wide Web in which automated software can store, exchange, and use machine-readable information distributed throughout the Web, in turn enabling users to deal with the information with greater efficiency and certainty.[4] RDF's simple data model and ability to model disparate, abstract concepts has also led to its increasing use in knowledge management applications unrelated to Semantic Web activity.

A collection of RDF statements intrinsically represents a labeled, directed multi-graph. As such, an RDF-based data model is more naturally suited to certain kinds of knowledge representation than the relational model and other ontological models. However, in practice, RDF data is often persisted in relational database or native representations also called Triplestores, or Quad stores if context (i.e. the named graph) is also persisted for each RDF triple.[5] As RDFS and OWL demonstrate, additional ontology languages can be built upon RDF.

[edit] History

There were several ancestors to the W3C's RDF. Technically the closest was MCF, a project initiated by Ramanathan V. Guha while at Apple Computer and continued, with contributions from Tim Bray, during his tenure at Netscape Communications Corporation. Ideas from the Dublin Core community, and from PICS, the Platform for Internet Content Selection (the W3C's early Web content labelling system) were also key in shaping the direction of the RDF project.

The W3C published a specification of RDF's data model and XML syntax as a Recommendation in 1999.[6] Work then began on a new version that was published as a set of related specifications in 2004. While there are a few implementations based on the 1999 Recommendation that have yet to be completely updated, adoption of the improved specifications has been rapid since they were developed in full public view, unlike some earlier technologies of the W3C. Most newcomers to RDF are unaware that the older specifications even exist.

In June 2010, W3C organized a workshop to gather feedback from the Web community and discuss possible revisions and improvements to RDF.[7]

Some libraries published their catalogue in RDF, one of them the Hungarian Széchényi Library.[8]

[edit] RDF Topics

[edit] RDF Vocabulary

The vocabulary defined by the RDF specification is as follows:[9]

[edit] Classes

[edit] rdf

- **`rdf:Resource`** - the class resource, everything
- **`rdf:XMLLiteral`** - the class of XML literal values
- **`rdf:Property`** - the class of properties
- **`rdf:Statement`** - the class of RDF statements
- **`rdf:Alt`**, **`rdf:Bag`**, **`rdf:Seq`** - containers of alternatives, unordered containers, and ordered containers (`rdfs:Container` is a super-class of the three)
- **`rdf:List`** - the class of RDF Lists
- **`rdf:nil`** - an instance of `rdf:List` representing the empty list

[edit] rdfs

- **`rdfs:Literal`** - the class of literal values, e.g. strings and integers
- **`rdfs:Class`** - the class of classes
- **`rdfs:Datatype`** - the class of RDF datatypes
- **`rdfs:Container`** - the class of RDF containers
- **`rdfs:ContainerMembershipProperty`** - the class of container membership properties, `rdf:_1`, `rdf:_2`, ..., all of which are sub-properties of `rdfs:member`

[edit] Properties

[edit] rdf

- **`rdf:type`** - an instance of `rdf:Property` used to state that a resource is an instance of a class
- **`rdf:first`** - the first item in the subject RDF list
- **`rdf:rest`** - the rest of the subject RDF list after `rdf:first`
- **`rdf:value`** - idiomatic property used for structured values
- **`rdf:subject`** - the subject of the subject RDF statement
- **`rdf:predicate`** - the predicate of the subject RDF statement
- **`rdf:object`** - the object of the subject RDF statement

`rdf:Statement`, `rdf:subject`, `rdf:predicate`, `rdf:object` are used for reification (see below).

[edit] rdfs

- **`rdfs:subClassOf`** - the subject is a subclass of a class
- **`rdfs:subPropertyOf`** - the subject is a subproperty of a property
- **`rdfs:domain`** - a domain of the subject property
- **`rdfs:range`** - a range of the subject property
- **`rdfs:label`** - a human-readable name for the subject
- **`rdfs:comment`** - a description of the subject resource
- **`rdfs:member`** - a member of the subject resource
- **`rdfs:seeAlso`** - further information about the subject resource

- **`rdfs:isDefinedBy`** - the definition of the subject resource

This vocabulary is used as a foundation for RDF Schema where it is extended.

[edit] Serialization formats

RDF/XML serialization

Filename extension	`.rdf`
Internet media type	`application/rdf+xml`[10]
Developed by	World Wide Web Consortium
Standard(s)	Concepts and Abstract Syntax February 10, 2004; 8 years ago
Open format?	Yes

Two common serialization formats are in use.

The first is an XML format. This format is often called simply RDF because it was introduced among the other W3C specifications defining RDF. However, it is important to distinguish the XML format from the abstract RDF model itself. Its MIME media type, application/rdf+xml, was registered by RFC 3870. It recommends RDF documents to follow the new 2004 specifications.

In addition to serializing RDF as XML, the W3C introduced Notation 3 (or N3) as a non-XML serialization of RDF models designed to be easier to write by hand, and in some cases easier to follow. Because it is based on a tabular notation, it makes the underlying triples encoded in the documents more easily recognizable compared to the XML serialization. N3 is closely related to the Turtle and N-Triples formats.

Triples may be stored in a triplestore.

[edit] Resource identification

The subject of an RDF statement is either a Uniform Resource Identifier (URI) or a blank node, both of which denote resources. Resources indicated by blank nodes are called anonymous resources. They are not directly identifiable from the RDF statement. The predicate is a URI which also indicates a resource, representing a relationship. The object is a URI, blank node or a Unicode string literal.

In Semantic Web applications, and in relatively popular applications of RDF like RSS and FOAF (Friend of a Friend), resources tend to be represented by URIs that intentionally denote, and can be used to access, actual data on the World Wide Web. But RDF, in general, is not limited to the description of Internet-based resources. In fact, the URI that names a resource does not have to be dereferenceable at all. For example, a URI that begins with "http:" and is used as the subject of an RDF statement does not necessarily have to represent a resource that is accessible via HTTP, nor does it need to represent a tangible, network-accessible resource — such a URI could represent absolutely anything. However, there is broad agreement that a bare URI (without a # symbol) which returns a 300-level coded response when used in an HTTP GET request should be treated as denoting the internet resource that it succeeds in accessing.

Therefore, producers and consumers of RDF statements must agree on the semantics of resource identifiers. Such agreement is not inherent to RDF itself, although there are some controlled vocabularies in common use, such as Dublin Core Metadata, which is partially mapped to a URI space for use in RDF. The intent of publishing RDF-based ontologies on the Web is often to establish, or circumscribe, the intended meanings of the resource identifiers used to express data in RDF. For example, the URI http://www.w3.org/TR/2004/REC-owl-guide-20040210/wine#merlot is intended by its owners to refer to the class of all Merlot red wines, an intent which is expressed by the OWL ontology — itself an RDF document — in which it occurs. Note that this is not a 'bare' resource identifier, but is rather a URI reference, containing the '#' character and ending with a fragment identifier.

[edit] Statement reification and context

The body of knowledge modeled by a collection of statements may be subjected to reification, in which each *statement* (that is each triple *subject-predicate-object* altogether) is assigned a URI and treated as a resource about which additional statements can be made, as in "*Jane says that* John is the author of document X". Reification is sometimes important in order to deduce a level of confidence or degree of usefulness for each statement.

In a reified RDF database, each original statement, being a resource, itself, most likely has at least three additional statements made about it: one to assert that its subject is some resource, one to assert that its predicate is some resource, and one to assert that its object is some resource or literal. More statements about the original statement may also exist, depending on the application's needs.

Borrowing from concepts available in logic (and as illustrated in graphical notations such as conceptual graphs and topic maps), some RDF model implementations acknowledge that it is sometimes useful to group statements according to different criteria, called *situations*, *contexts*, or *scopes*, as discussed in articles by RDF specification co-editor Graham Klyne.[11][12] For example, a statement can be associated with a context, named by a URI, in order to assert an "is true in" relationship. As another example, it is sometimes convenient to group statements by their source, which can be identified by a URI, such as the URI of a particular RDF/XML document. Then, when updates are made to the source, corresponding statements can be changed in the model, as well.

Implementation of scopes does not necessarily require fully reified statements. Some implementations allow a single scope identifier to be associated with a statement that has not been assigned a URI, itself.[13][14] Likewise *named graphs* in which a set of triples is named by a URI can represent context without the need to reify the triples.[15]

[edit] Query and inference languages

The predominant query language for RDF graphs is SPARQL. SPARQL is an SQL-like language, and a recommendation of the W3C as of January 15, 2008.

An example of a SPARQL query to show country capitals in Africa, using a fictional ontology.

```
PREFIX abc: <nul://sparql/exampleOntology#> .
SELECT ?capital ?country
WHERE {
  ?x abc:cityname ?capital ;
     abc:isCapitalOf ?y.
  ?y abc:countryname ?country ;
     abc:isInContinent abc:Africa.
}
```

Other ways to query RDF graphs include:

- RDQL, precursor to SPARQL, SQL-like
- Versa, compact syntax (non–SQL-like), solely implemented in 4Suite (Python)
- RQL, one of the first declarative languages for uniformly querying RDF schemas and resource descriptions, implemented in RDFSuite.[16]
- SeRQL, part of Sesame
- XUL has a template element in which to declare rules for matching data in RDF. XUL uses RDF extensively for databinding.

[edit] Examples

[edit] Example 1: RDF Description of a person named Eric Miller[17]

The following example is taken from the W3C website[17] describing a resource with statements "there is a Person identified by http://www.w3.org/People/EM/contact#me, whose name is Eric Miller, whose email address is em@w3.org, and whose title is Dr.".

An RDF Graph Describing Eric Miller[17]

The resource "http://www.w3.org/People/EM/contact#me" is the subject.

The objects are:

- "Eric Miller" (with a predicate "whose name is"),
- em@w3.org (with a predicate "whose email address is"), and
- "Dr." (with a predicate "whose title is").

The subject is a URI.

The predicates also have URIs. For example, the URI for each predicate:

- "whose name is" is http://www.w3.org/2000/10/swap/pim/contact#fullName,
- "whose email address is" is http://www.w3.org/2000/10/swap/pim/contact#mailbox,
- "whose title is" is http://www.w3.org/2000/10/swap/pim/contact#personalTitle.

In addition, the subject has a type (with URI http://www.w3.org/1999/02/22-rdf-syntax-ns#type), which is person (with URI http://www.w3.org/2000/10/swap/pim/contact#Person), and a mailbox (with URI http://www.w3.org/2000/10/swap/pim/contact#mailbox.)

Therefore, the following "subject, predicate, object" RDF triples can be expressed:

- http://www.w3.org/People/EM/contact#me, http://www.w3.org/2000/10/swap/pim/contact#fullName, "Eric Miller"
- http://www.w3.org/People/EM/contact#me, http://www.w3.org/2000/10/swap/pim/contact#personalTitle, "Dr."
- http://www.w3.org/People/EM/contact#me, http://www.w3.org/1999/02/22-rdf-syntax-ns#type, http://www.w3.org/2000/10/swap/pim/contact#Person
- http://www.w3.org/People/EM/contact#me, http://www.w3.org/2000/10/swap/pim/contact#mailbox, em@w3.org

[edit] Example 2: The postal abbreviation for New York

Certain concepts in RDF are taken from logic and linguistics, where subject-predicate and subject-predicate-object structures have meanings similar to, yet distinct from, the uses of those terms in RDF. This example demonstrates:

In the English language statement *'New York has the postal abbreviation NY'* , *'New York'* would be the subject, *'has the postal abbreviation'* the predicate and *'NY'* the object.

Encoded as an RDF triple, the subject and predicate would have to be resources named by URIs. The object could be a resource or literal element. For example, in the Notation 3 form of RDF, the statement might look like:

```
<urn:x-states:New%20York> <http://purl.org/dc/terms/alternative> "NY" .
```

In this example, "urn:x-states:New%20York" is the URI for a resource that denotes the U.S. state New York, "http://purl.org/dc/terms/alternative" is the URI for a predicate (whose human-readable definition can be found at here [18]), and "NY" is a literal string. Note that the URIs chosen here are not standard, and don't need to be, as long as their meaning is known to whatever is reading them.

N-Triples is just one of several standard serialization formats for RDF. The triple above can also be equivalently represented in the standard RDF/XML format as:

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dcterms="http://purl.org/dc/terms/">
        <rdf:Description rdf:about="urn:x-states:New%20York">
                        <dcterms:alternative>NY</dcterms:alternative>
        </rdf:Description>
</rdf:RDF>
```

However, because of the restrictions on the syntax of QNames (such as `dcterms:alternative` above), there are some RDF graphs that are not representable with RDF/XML.

[edit] Example 3: A Wikipedia article about Tony Benn

In a like manner, given that "http://en.wikipedia.org/wiki/Tony_Benn" identifies a particular resource (regardless of whether that URI could be traversed as a hyperlink, or whether the resource is *actually* the Wikipedia article about Tony Benn), to say that the title of this resource is "Tony Benn" and its publisher is "Wikipedia" would be two assertions that could be expressed as valid RDF statements. In the N-Triples form of RDF, these statements might look like the following:

```
<http://en.wikipedia.org/wiki/Tony_Benn> <http://purl.org/dc/elements/1.1/title> "Tony
```

```
Benn" .
<http://en.wikipedia.org/wiki/Tony_Benn> <http://purl.org/dc/elements/1.1/publisher>
"Wikipedia" .
```

And these statements might be expressed in RDF/XML as:

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
        <rdf:Description rdf:about="http://en.wikipedia.org/wiki/Tony_Benn">
                <dc:title>Tony Benn</dc:title>
                <dc:publisher>Wikipedia</dc:publisher>
        </rdf:Description>
</rdf:RDF>
```

To an English-speaking person, the same information could be represented simply as:

> The title of this resource, which is published by Wikipedia, is 'Tony Benn'

However, RDF puts the information in a formal way that a machine can understand. The purpose of RDF is to provide an encoding and interpretation mechanism so that resources can be described in a way that particular software can understand it; in other words, so that software can access and use information that it otherwise couldn't use.

Both versions of the statements above are wordy because one requirement for an RDF resource (as a subject or a predicate) is that it be unique. The subject resource must be unique in an attempt to pinpoint the exact resource being described. The predicate needs to be unique in order to reduce the chance that the idea of Title or Publisher will be ambiguous to software working with the description. If the software recognizes *http://purl.org/dc/elements/1.1/title* (a specific definition for the concept of a title established by the Dublin Core Metadata Initiative), it will also know that this title is different from a land title or an honorary title or just the letters t-i-t-l-e put together.

The following example shows how such simple claims can be elaborated on, by combining multiple RDF vocabularies. Here, we note that the primary topic of the Wikipedia page is a "Person" whose name is "Tony Benn":

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
        <rdf:Description rdf:about="http://en.wikipedia.org/wiki/Tony_Benn">
                <dc:title>Tony Benn</dc:title>
                <dc:publisher>Wikipedia</dc:publisher>
                <foaf:primaryTopic>
                     <foaf:Person>
                          <foaf:name>Tony Benn</foaf:name>
                     </foaf:Person>
                </foaf:primaryTopic>
        </rdf:Description>
</rdf:RDF>
```

[edit] Applications

- Sigma - Application from DERI in National University of Ireland, Galway(NUIG).

- Creative Commons - Uses RDF to embed license information in web pages and mp3 files.
- DOAC (Description of a Career) - supplements FOAF to allow the sharing of résumé information.
- FOAF (Friend of a Friend) - designed to describe people, their interests and interconnections.
- Haystack client - Semantic web browser from MIT CS & AI lab.[19]
- IDEAS Group - developing a formal 4D Ontology for Enterprise Architecture using RDF as the encoding.[20]
- Microsoft shipped a product, Connected Services Framework,[21] which provides RDF-based Profile Management capabilities.
- MusicBrainz - Publishes information about Music Albums.[22]
- NEPOMUK, an open-source software specification for a Social Semantic desktop uses RDF as a storage format for collected metadata. NEPOMUK is mostly known because of its integration into the KDE SC 4 desktop environment.
- RDF Site Summary - one of several "RSS" languages for publishing information about updates made to a web page; it is often used for disseminating news article summaries and sharing weblog content.
- ResumeRDF - developed to express information contained in a personal Resume or Curriculum Vitae (CV) on the Semantic Web. This includes information about work and academic experience, skills, etc.
- Simple Knowledge Organization System (SKOS) - a KR representation intended to support vocabulary/thesaurus applications
- SIOC (Semantically-Interlinked Online Communities) - designed to describe online communities and to create connections between Internet-based discussions from message boards, weblogs and mailing lists.[23]
- Smart-M3 - provides an infrastructure for using RDF and specifically uses the ontology agnostic nature of RDF to enable heterogeneous mashing-up of information[24]
- Many other RDF schemas are available by searching SchemaWeb.[25]

Some uses of RDF include research into social networking. This is important because it could help governments keep track of undesirables. It will also help people in business fields understand better their relationships with members of industries that could be of use for product placement.[26] It will also help scientists understand how people are connected to one another.

RDF is being used to have a better understanding of traffic patterns. This is because the information regarding traffic patterns is on different websites, and RDF is used to integrate information from different sources on the web. Before, the common methodology was using keyword searching, but this method is problematic because it does not consider synonyms. This is why ontologies are useful in this situation. But one of the issues that comes up when trying to efficiently study traffic is that to fully understand traffic, concepts related to people, streets, and roads must be well understood. Since these are human concepts, they require the addition of fuzzy logic. This is because values that are useful when describing roads, like slipperiness, are not precise concepts and cannot be measured. This would imply that the best solution would incorporate both fuzzy logic and ontology.[27]

[edit]