

General area: Advanced Topics in Databases (CS848)

Course WEB page: <http://cs.uwaterloo.ca/~gweddell/cs848>

Instructor: Grant Weddell (email: gweddell@uwaterloo.ca)

Location and time: TBD (scheduled for Winter, 2015)

Organization: This is a seminar course consisting of presentations by the instructor and paper presentations by the students.

Data Access on the WEB

Consider the HTML associated with the URI

http://en.wikipedia.org/wiki/Resource_Description_Framework

To humans (with browser):

The screenshot shows a web browser window displaying the Wikipedia page for "Resource Description Framework". The browser's address bar shows the URL: http://en.wikipedia.org/wiki/Resource_Description_Framework. The page content is as follows:

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 notations and data serialization formats.

Contents [hide]

- 1 Overview
- 2 History
- 3 RDF topics
 - 3.1 RDF vocabulary
 - 3.1.1 Classes
 - 3.1.1.1 rdf
 - 3.1.1.2 rdfs
 - 3.1.2 Properties
 - 3.1.2.1 rdf
 - 3.1.2.2 rdfs
 - 3.2 Serialization formats
 - 3.3 Resource identification
 - 3.4 Statement replication and context
 - 3.5 Query and inference languages
- 4 Examples
 - 4.1 Example 1: RDF Description of a person named Eric Miller^[2]
 - 4.2 Example 2: The postal abbreviation for New York
 - 4.3 Example 3: A Wikipedia article about Tony Benn
- 5 Applications
- 6 See also
- 7 References
- 8 Further reading
- 9 External links

Overview [edit source | edit meta]

The RDF data model^[3] 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 sentences. These sentences are known as **triple RDF sentences**. They don't create the resources, and the predicate

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 ⓘ

Data Access on the WEB

Consider the HTML associated with the URI

http://en.wikipedia.org/wiki/Resource_Description_Framework

To machines:



```
</li>
<li class="toclevel-1 toctection-19"><a href="#">Applications</a></li>
<li class="toclevel-1 toctection-20"><a href="#">See also</a></li>
<li class="toclevel-1 toctection-21"><a href="#">References</a></li>
<li class="toclevel-1 toctection-22"><a href="#">Further reading</a></li>
<li class="toclevel-1 toctection-23"><a href="#">External links</a></li>
</ul>
</div>
<h2>Overview</h2>
<div class="mw-editsection">
</div>
<p>The RDF data model<sup>[cite ref-2]</sup> class="reference"> is an abstract model of making <sup>[cite note-2]</sup> relationships between entities. It is similar to classic conceptual modeling approaches such as <sup>[cite note-3]</sup> relationship model<sup>[cite ref-2]</sup> class="reference">, <sup>[cite note-2]</sup> class="reference"> class diagrams<sup>[cite ref-2]</sup> class="reference">, and <sup>[cite note-4]</sup> class="reference"> state transition diagrams<sup>[cite ref-2]</sup> class="reference">. It is based upon the idea of making <sup>[cite note-2]</sup> class="reference"> statements about <sup>[cite note-2]</sup> class="reference"> resources<sup>[cite ref-2]</sup> class="reference"> in the form of subject-predicate-object expressions. These expressions are known as <sup>[cite note-2]</sup> class="reference"> triples<sup>[cite ref-2]</sup> class="reference"> 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 <sup>[cite note-2]</sup> class="reference"> subject (grammar)<sup>[cite ref-2]</sup> class="reference"> denoting "the sky", a <sup>[cite note-2]</sup> class="reference"> predicate (grammar)<sup>[cite ref-2]</sup> class="reference"> denoting "has the color", and a <sup>[cite note-2]</sup> class="reference"> object (grammar)<sup>[cite ref-2]</sup> class="reference"> denoting "blue". Therefore RDF swaps object for subject that would be used in the classical notation of an <sup>[cite note-2]</sup> class="reference"> entity-attribute-value model<sup>[cite ref-2]</sup> class="reference"> within <sup>[cite note-2]</sup> class="reference"> object-oriented design<sup>[cite ref-2]</sup> class="reference">. RDF is an abstract model with several <sup>[cite note-2]</sup> class="reference"> serialization formats<sup>[cite ref-2]</sup> class="reference"> (i.e., file formats), and so the particular way in which a resource or triple is encoded varies from format to format.<sup>[cite ref-3]</sup> class="reference">
```

Consider the HTML associated with the URI

http://en.wikipedia.org/wiki/Resource_Description_Framework

To usefully integrate this information, machines must

1. understand natural languages, and
2. **have domain specific understanding of the world.**

RDF, OWL and SPARQL are W3C recommendations on addressing the **second** problem:

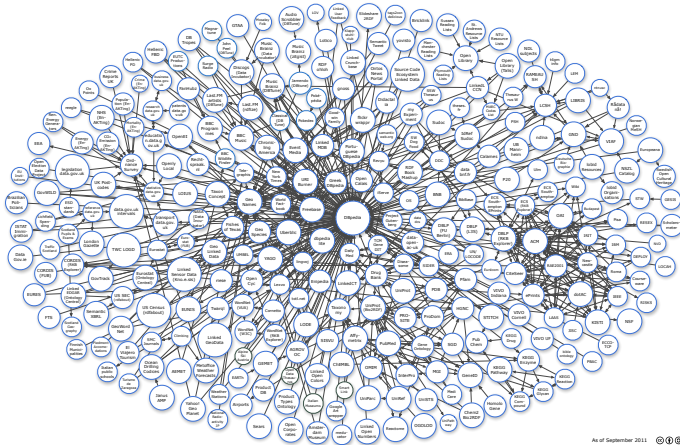
1. Add HTML that encodes *data* and *metadata* in the form of RDF *subject/predicate/object* triples and
2. Add a **new** WEB function to extract the RDF triples from HTML

Fetch : URI \rightarrow HTML

SelectRDF : HTML \rightarrow RDF

Data Access on the WEB

There is now lots of RDF data and metadata!



As of September 2011 © 1 2

Linking Open Data cloud diagram, by Richard Cyganiak and Anja Jentzsch.
<http://lod-cloud.net/>

Data Access on the WEB

Example: `SelectRDF(Fetch("http://www.acme.ca/emp-knowledge"))` returns following RDF

<i>(data)</i>	<i>(metadata)</i>
John/age/32	student/subclass/human
John/type/student	human/exists-property/age
Mary/type/student	age/range/integer
	age/type/functional-property

The query *known humans and their ages* is given in SPARQL as

```
select ?x, ?y
where {?x type human. ?x age ?y}
```

When evaluated over the RDF, returns

- ▶ With *basic RDF* entailment: { }
- ▶ With *OWL 2* entailment: {{?x = John, ?y = 32}}

Data Access on the WEB

Example: `SelectRDF(Fetch("http://www.acme.ca/emp-knowledge"))` returns following RDF

<i>(data)</i>	<i>(metadata)</i>
John/age/32	student/subclass/human
John/type/student	human/exists-property/age
Mary/type/student	age/range/integer
	age/type/functional-property

The query *known humans that have an age* is given in SPARQL as

```
select ?x
where {?x type concept-intersection(human, exists-property(age))}
```

Relies on OWL 2's ability to express **complex concepts**: *the intersection of the set of humans and the set of things having an age property*

When evaluated over the RDF, returns

- ▶ With *basic RDF* entailment: { }
- ▶ With *OWL 2* entailment: {{?x = John}, {?x = Mary}}

There are three main topics:

1. RDF storage engines,
2. SPARQL query optimization and evaluation and
3. *Ontology-Based Data Access* (OBDA).

Will also be useful to review

- ▶ complexity theory, to understand the difficulty of query evaluation,
- ▶ first order logic, which underlies both the relational and WEB setting and
- ▶ *description logics* (DLs), in particular the dialect *SHROIQ(D)*.

Big topics that are currently beyond the scope of the course:

1. *fact extraction* (raw html to RDF data),
2. *schema integration* (combining RDF metadata) and
3. *inductive reasoning and schema evolution* (raw html to RDF metadata).