Data Access on the WEB: Topics on RDF, SPARQL 1.1 and OWL 2

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)

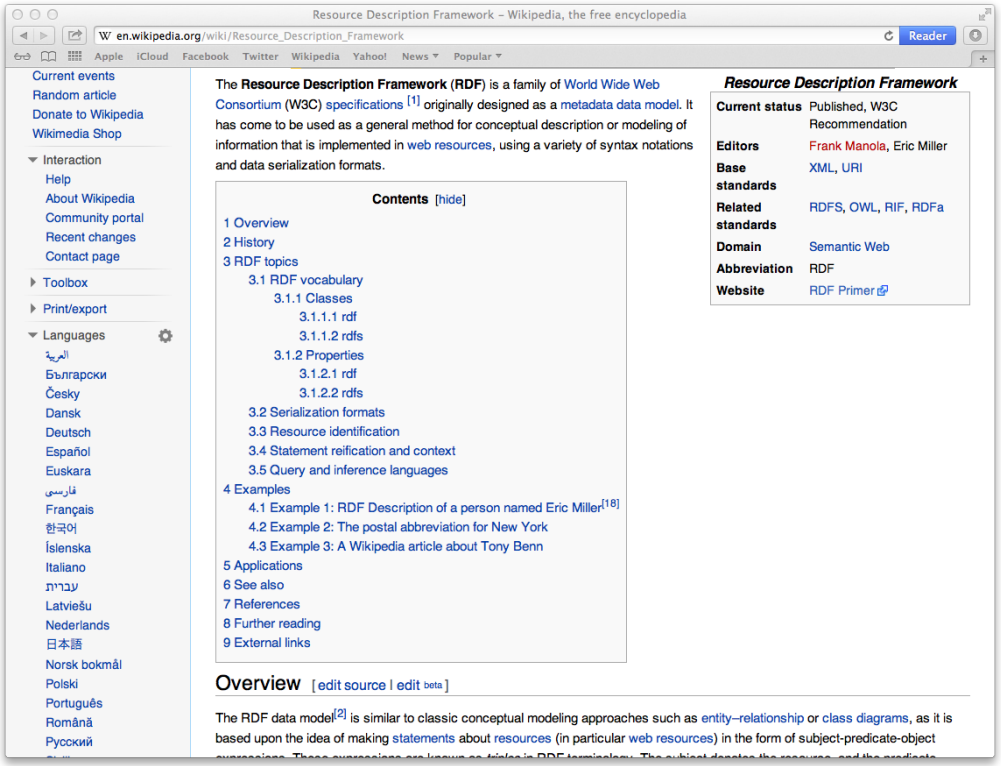
Organization: This is a seminar course consisting of presentations by the instructor and paper presentations by the students.

Data Access on the WEB

Consider the HTML associated with the URI

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

To humans (with browser):



Data Access on the WEB

Consider the HTML associated with the URI

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

To machines:

Data Access on the WEB

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.

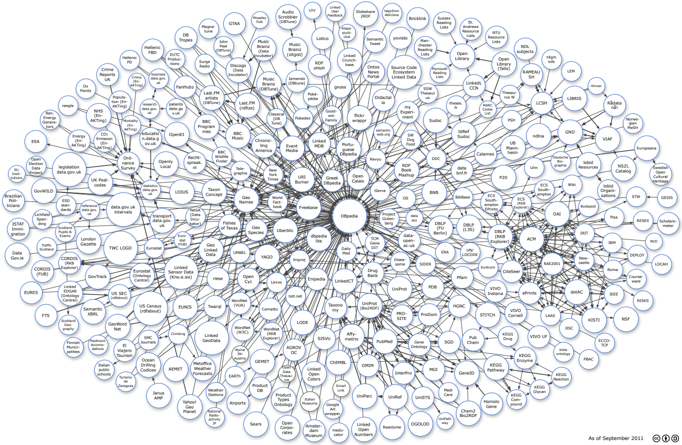
RDF, OWL and SPARQL are W3C recommendations on addressing the second problem:

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

$$\text{Fetch} \quad : \quad \text{URI} \rightarrow \text{HTML}$$

$$\text{SelectRDF} \quad : \quad \text{HTML} \rightarrow \text{RDF}$$

There is now lots of RDF data and metadata!



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

(*data*)	(*metadata*)
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: $\{\{?x = \text{John}, ?y = 32\}\}$

Data Access on the WEB

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

(*data*)	(*metadata*)
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}}

Overview

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 $\mathcal{SHROIQ}(\mathbf{D})$.

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).