COMP318: RDF serialisations

www.csc.liv.ac.uk/~valli/Comp318

Dr Valentina Tamma

Room: Ashton 2.12

Dept of computer science

University of Liverpool

V.Tamma@liverpool.ac.uk

Where were we

- RDF:
  - data model for sharing information on the Web
  - syntax
  - triples
- RDFS
  - language for defining the schema representing a domain of interest
  - classes and properties
- RDF vocabulary
  - the constructs we use to represent data and knowledge in RDF/S
- RDFS semantics for inferring new information from the graph
- OWL syntax and modelling primitives

Why serialisation?

- RDF is not a data format, but a **data model** for describing resources in the forms of triples
  - *subject, predicate, object*

- One of the use cases underlying RDF's design is to describe (***mark up***) the content of HTML web pages.
  - RDFa is a syntactic variant introduced to help with that use case
  - RDFa embeds RDF within the attributes of HTML tags

Why serialisation?

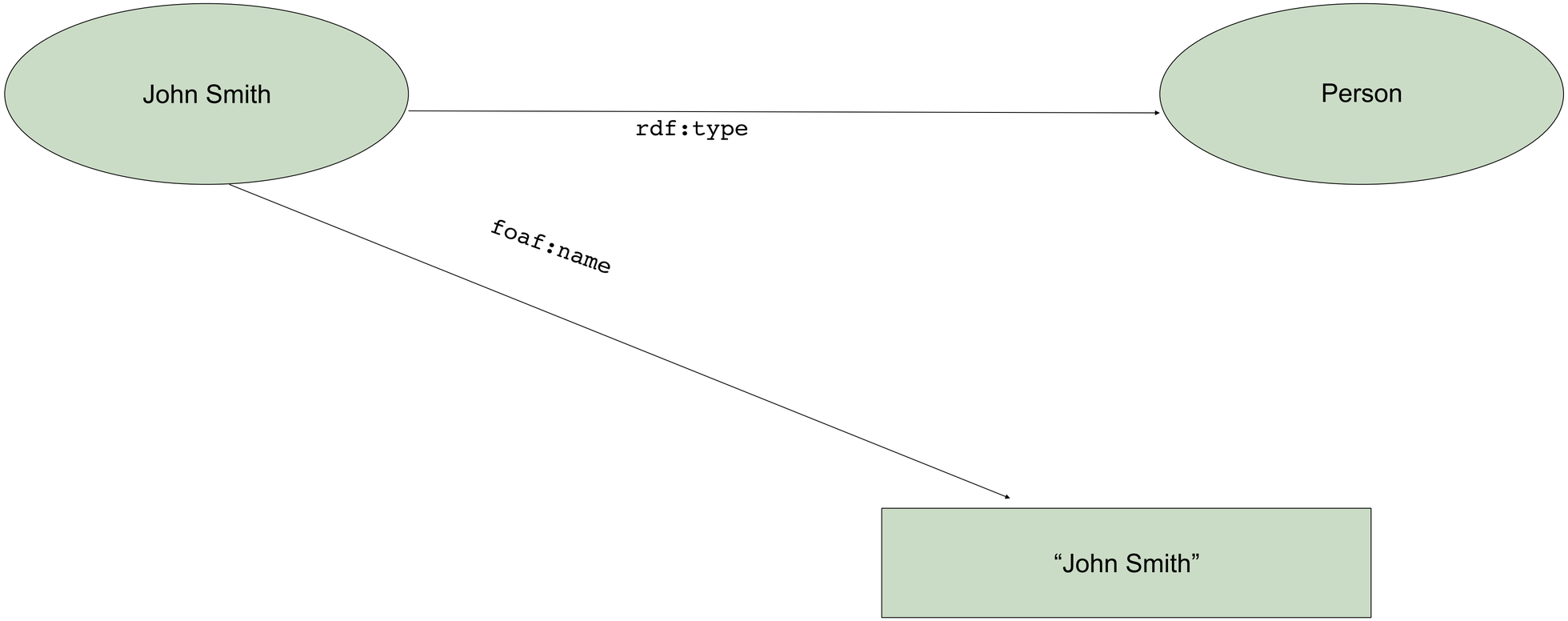
- RDF is not a data format, but a **data model** for describing resources in the forms of triples
  - *subject, predicate, object*

- To publish a document on the Web it must be serialised using one of the RDF syntax formats
  - the triples forming the graph representing the RDF document need to be written in a file
    - conforming to the syntactic constraints deriving from the chosen syntax
      - in advance if dealing with static data set
      - on demand if dealing with dynamic data sets

Serialisation formats

- RDF serialisatio formalisms
  - RDF/XML
  - RDFa
  - Turtle
  - N-Triples
  - RDF/JSON

A simple example



RDF/XML

- The **`MIME`** type for the document
  - indicates the Internet media type and indicates the format of files on the Internet
  - for RDF/XML documents is **`application/rdf+xml`**

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
 xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
 xmlns:foaf=”http://xmlns.com/foaf/0.1/”>

 <rdf:Description rdf:about="http://www.liv.ac.uk/staff/john-smith">
 <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Person”/>
   <foaf:name>John Smith</foaf:name>
 </rdf:Description>
</rdf:RDF>
```

RDF/XML explained

- The example shows two triples
  - The resource identified by the URI `http://www.liv.ac.uk/staff/john-smith` is of type `Person`
    - `Person` is defined in the FOAF vocabulary
    - the URI identifies a document that is a reference to John Smith
  - The `name` of the resource is “John Smith”
    - `name` is defined in the FOAF vocabulary

Turtle

- Turtle is the plain text serialisation format for RDF
  - typically used to edit RDF documents by hand
  - the **`MIME`** type for Turtle is **`text/turtle;charset=utf-8.`**

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@ prefix foaf: <http://xmlns.com/foaf/0.1/> .

<http://www.liv.ac.uk/staff/john-smith>
   rdf:type foaf:Person;
   foaf:name “John Smith” .
```

N-Triples

```
<http://www.liv.ac.uk/staff/john-smith>
   <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
   <http://xmlns.com/foaf/0.1/Person> .

<http://www.liv.ac.uk/staff/john-smith>
   <http://xmlns.com/foaf/0.1/name>
   “John Smith” .
```

N-Triples

- N-Triples is a subset of Turtle
  - it does not allow namespace prefixes and shorthand
  - introduces redundancy in the serialisation format
    - allows N-Triples files to be parsed one line at the time
      - no memory constraints over large files
    - N-Triples files can be easily compressed
      - reduction in network traffic when exchanging files
    - currently accepted standard for exchanging large dumps of Linked Data

RDF/JSON

- JSON (JavaScript Object Notation) serialisation for RDF
  - highly desirable as many programming languages provide JSON support
    - triples as nested data structures
  - it represents simple data structures and associative arrays:
    - `Number`
    - `String`
    - `Boolean`
    - `Array`: an ordered sequence of values, comma-separated and enclosed in square brackets; the values do not need to be of the same type.
    - `Object:` an unordered collection of key:value pairs with the ':' character separating the key and the value, comma-separated and enclosed in curly braces;
    - `null`
  - can facilitate uptake of mobile applications consuming semantic data
    - `MIME` content-type is `application/json`

RDF/JSON

```
{
   "prefixes" : {
       “rdf” : http://www.w3.org/1999/02/22-rdf-syntax-ns#
       "foaf" : "http://xmlns.com/foaf/0.1/" ,
   } ,

   <http://www.liv.ac.uk/staff/john-smith> : {
       "rdf:type" : [
           { "value" : "foaf:Person", "type" : "rdf:Resource" }
       ],
       "http://xmlns.com/foaf/0.1/name" : [
           { "value" : "John Smith", "type" : "literal" }
       ]
    }
}
```

Subject

Predicate

Object

RDF/JSON

- Example explained:
  - data represented in dictionaries:
    - two elements in a dictionary, with the keys: "**`prefixes`**" and “**`http://www.liv.ac.uk/staff/john_smith`**“
    - Their values are both dictionaries

RDFa

- RDFa is a W3C recommendation that adds a set of attribute level extensions to X(HTML)/XML for embedding rich metadata.
  - it supports the notions of namespaces and URIs
    - it allows the mixing of vocabularies as in RDF
  - it offers a flexible framework for using Resources of type URI or Literal
  - it is a complete serialisation of RDF
    - `http://www.w3.org/TR/rdfa-syntax/`

RDFa: embedding in HTML

- RDFa (RDF attribute) allows the embedding of semantic information in existing (X)HTML documents
  - extends (X)HTML a bit by:
    - defining general attributes to add metadata to any elements
    - provides an almost complete “serialisation” of RDF in XHTM
  - designed to enrich existing pages that have already limited semantics based on hyperlinks and tag layout
    - `div` and `span`
  - the RDF data is embedded within the HTML DOM
    - existing content can be annotated with RDFa by simply modifying the HTML document

Main principles of RDFa

- RDFa means “RDF in attributes”:
  - all RDF contents are defined through XML **attributes**
    - no elements
  - it uses the XML/HTML tree structure
  - many of the attributes are defined by RDFa
    - some attributes (`@href`, `@rel`) are also reused
    - `@href` and `@rel` represent objects that are URI references
  - the text content (`@content`) is also reused for literals as well as `@href` values

RDFa

```
<!DOCTYPE html PUBLIC "-//W3C//DTDXHTML+RDFa 1.0//EN"
   “http://www.w3.org/MarkUp/DTD/xhtml-rdfa-1.dtd”>
<html xmlns=”http://www.w3.org/1999/xhtml”
   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
   xmlns:foaf=”http://xmlns.com/foaf/0.1/”>

<head>
   <meta http-equiv=”Content-Type" content="application/xhtml+xml;
      charset=”UTF-8”/>
   <title>John Smith’s personal page</title>
</head>

<body>
   <div about=”http://liv.ac.uk/staff#john-smith” typeof=”foaf:Person”
      <span property="foaf:name">John Smith</span>
   </div>
</body>

</html>
```

Main principles of RDFa

RDFa is a serialization of RDF embedded in XHTML, HTML, or XML in general

- Most of the data on the web are in (X)HTML:
  - new content generated every day
  - how do we get structured data from that info?

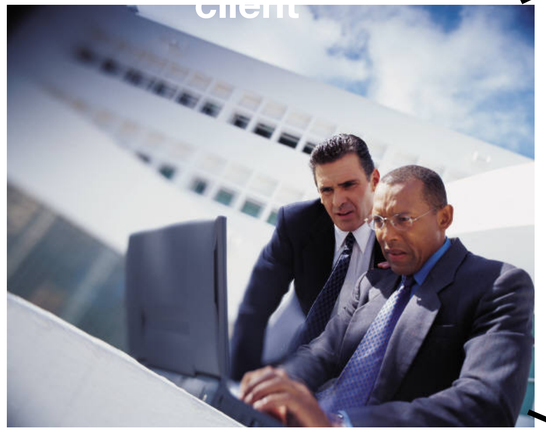
- Especially when authors of the “traditional web” don’t like to generate RDF/XML files separately
  - RDF/XML is complex
  - it requires a separate storage, generation, etc. mechanism
    - that is also valid for, e.g., Turtle
    - but even when authoring with a text editor, creating an extra file is a load

What does this mean in practice?

- The same (X)HTML file:
  - is used, unchanged, by browsers
    - they ignore attributes they do not know
  - can be used by specialised processors (or APIs) to extract RDF triples

Typical usage pattern

Request for
`http://www.w3.org/ns/entailment/data/RDFS`



Request for
`http://www.w3.org/ns/entailment/data/RDFS.ttl`

http://www.w3.org/ns/entailment/data/RDFS.html

W3C Semantic Web

Unique identifier for *RDFS Entailment*.

"http://www.w3.org/ns/entailment/RDFS" is the URI. The specification for the RDFS entailment is part of the RDF Semantics W3C Recommendation.

For more information about RDF, please refer to the the RDF Concepts and Abstract Syntax Recommendation .

Ivan Herman, ivan@w3.org, W3C, Semantic Web Activity Lead, 2009-05-03

http://www.w3.org/ns/entailment/data/RDFS.ttl

```
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .


ent:RDFS a ent:Entailment ;
     dc:creator <http://www.ivan-herman.net/foaf#me> ;
     dc:date "2009-05-03" ;
     dc:description "Unique identifier for RDFS Entailment" ;
     rdfs:comment "The specification for the RDFS entailment is part of the RDF Semantics
W3C Recommendation." ;
     rdfs:isDefinedBy <http://www.w3.org/TR/2004/REC-rdf-mt-20040210/#rdfs_entailment> ;
     rdfs:seeAlso <http://www.w3.org/TR/2004/REC-rdf-mt-20040210/> .

<http://www.w3.org/ns/entailment/data/RDFS.html> dc:title "Information Resource RDFS
Entailment" ;
     xhv:stylesheet <http://www.w3.org/StyleSheets/TR/base> .
```

RDFa

- adds new (X)HTML/XML attributes
- has namespaces and URI-s at its core; i.e., mixing vocabulary is just as easy as in RDF
- complete flexibility for using Literals or URI Resources
  - is a complete serialization of RDF

- RDFa is a bridge between the Web of Documents and the Web of Data

Where does the Turtle content come from?

- The triples are embedded in the HTML file
  - a client may know how to extract RDF triples directly from that file; or
  - an online “distiller” service is used; or
  - the server is set up to generate the Turtle file automatically

- However... the content is created only once!

Recap

- Serialisation:
  - RDF document published in a chosen syntax
  - XML/RDF, Turtle and N-triples, **RDFa**, RDF/JSON