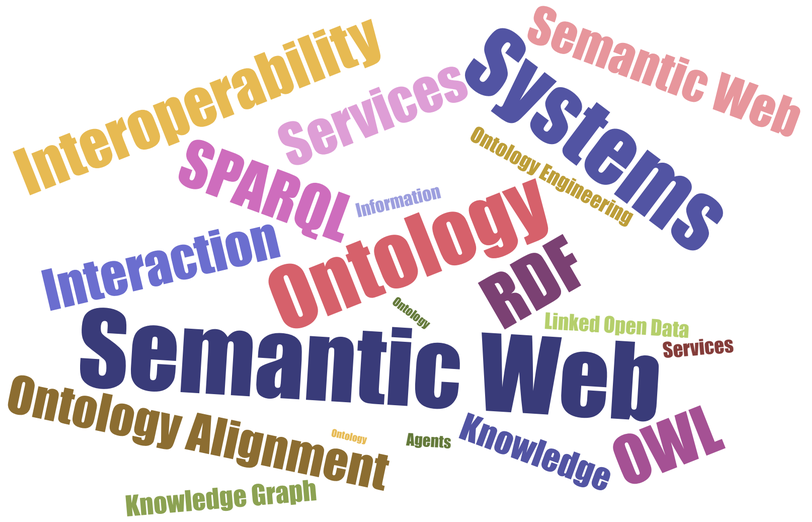
COMP318

Ontologies and Semantic Web



RDF - Part 6

Dr Valentina Tamma

V.Tamma@liverpool.ac.uk

Where were we

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

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

Baron Way apartment for sale

- Pure HTML page
  - No machine readable description

```
<html>
   <body>
    <H1> Baron Way Apartment for Sale</H1>
       The Baron Way Apartment has three bedrooms and is located in
       the family friendly Baron Way Building. The Apartment is located
       in the north of Amsterdam.
   </body>
</html>
```

Baron Way apartment for sale

- Pure HTML page
  - No machine readable description

```
<html>
   <body>
    <H1> Baron Way Apartment for Sale</H1>
       The Baron Way Apartment has three bedrooms and is located in
       the family friendly Baron Way Building. The Apartment is located
       in the north of Amsterdam.
   </body>
</html>
```

Namespace declaration

- Embedding of semantic information in existing XHTML documents
  - Exploits the **`xmlns`** attribute to define the namespaces to be used in the annotation
  - Similar to RDF/XML

```
<html
   xmlns:dbpedia="http://dbpedia.org/resource/"
   xmlns:dbpediaowl="http://dbpedia.org/ontology/"
   xmlns:swp="http://www.swpExample.org/ontology/flats.ttl#"
    xmlns:geo="http://www.geonames.org/ontology#">
```

Triple annotation

- RDF triples are encoded in tags such as spans, paragraphs and links, so they are not rendered when the browser displays the HTML code
  - Subjects are identified in the **`about`** attribute
  - Statements where the object is a literal are identified by the HTML attribute **`property`**

```
<body>
<H1> Baron Way Apartment for Sale</H1>
<div about="[swp:BaronWayFlat]">
The Baron Way Flat has
<span property="swp:hasNumberOfBedrooms">3</span> bedrooms and is
located in the family friendlyBaron Way Building</span>
```

Triple annotation

- RDF triples are encoded in tags such as spans, paragraphs and links, so they are not rendered when the browser displays the HTML code
  - Statements where the object is a resource are identified by the HTML attribute **`rel`**
  - Followed by the attribute **`resource`**

```
<span rel="swp:isPartOf"
  resource=“[swp:BaronWayBuilding]"> Baron Way Building</span>
<div about="[swp:BaronWayBuilding]">
The building is located in the north of Amsterdam. <span
rel="dbpediaowl:location"
     resource="[dbpedia:Amsterdam]"></span>
```

The entire code

```
<html
   xmlns:dbpedia="http://dbpedia.org/resource/"
   xmlns:dbpediaowl="http://dbpedia.org/ontology/"
   xmlns:swp="http://www.swpExample.org/ontology/flats.ttl#"
    xmlns:geo="http://www.geonames.org/ontology#">
<body>
<H1> Baron Way Apartment for Sale</H1>
<div about="[swp:BaronWayFlat]">
The Baron Way Flat has
<span property="swp:hasNumberOfBedrooms">3</span> bedrooms and is located in the family
friendlyBaron Way Building</span><span rel="swp:isPartOf"
   resource="[swp:BaronWayBuilding]"> Baron Way Building</span>
<div about="[swp:BaronWayBuilding]">
The building is located in the north of Amsterdam. <span rel="dbpediaowl:location"
     resource="[dbpedia:Amsterdam]"></span>
    <span rel="dbpediaowl:location"
     resource="[dbpedia:Netherlands]"></span>
  </div>
 </div> </body> </html>
```

The Turtle in RDFa

- Is that it?
  - The combination of **`@about`** with **`@rel`** / **`@property`** and possibly **`@href`** covers most of what we need...
  - but this is too complex for authors
- Go Turtle:
  - Use compact URIs when possible
  - Make use of the natural structure for
    - shared subjects
    - shared predicates
    - create blank nodes
    - ...

RDFa supported attributes

- `xmlns`: a prefix and qualified URL defining a namespace for a document;
- `about`: a resource URI or CURIE used to represent the subject in an RDF triple
- `property`: a white-space separated list of CURIEs representing predicates between a subject and a plain literal.
- `rel`: represents predicates between a subject and another resource
  - No literals!
  - CURIES can be considered a datatype found both in XML and non-XML grammars
    - syntax: [isbn:0393315703]
- `rev`: similar to rel, but traverses the predicate in the opposite direction wrt `rel`
- `http://www.w3.org/TR/xhtml-rdfa-primer/`

Which one

- Use N-Triples / N-Quads if you want decent performance and high compatibility.
- Use JSON / JSON LD if you want to improve your exsting JSON API, and don’t need performant RDF parsing.
- Use Turtle if you want to manually read & edit your RDF.
- Use Notation3 if you need RDF rules.
- Use RDFa to extend your existing HTML pages.
- Use RDF/XML if you need to use XML.
- If you can, support all of them

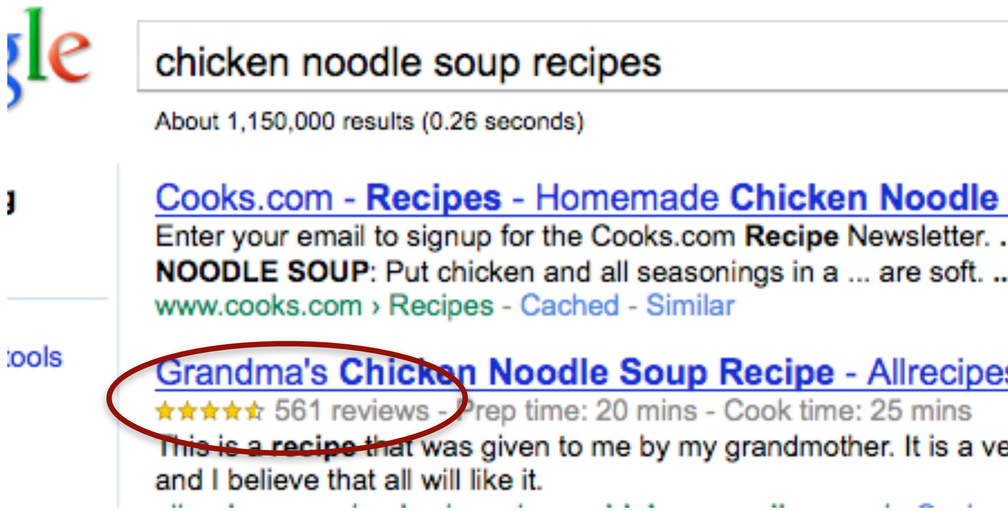
Consuming RDFa

- Various search engines begin to consume RDFa
  - Google, Yahoo, …
    - they may specify which vocabularies they “understand”
    - this is still an evolving area

- Facebook’s “social graph” is based on RDFa

Google’s rich sniplet

- Embedded metadata (microformat or RDFa) is used to improve search result page
  - at the moment only a few vocabularies are recognised, but that will evolve over the years

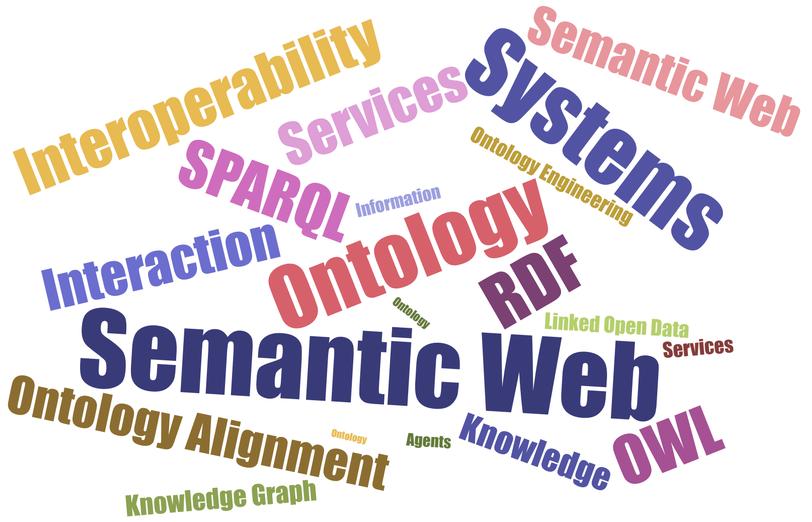


Who uses it

- A number of popular sites publish RDFa as part of their normal pages:
  - Google.com
  - Youtube.com
  - Facebook.com
  - Wikipedia.org
  - Yahoo.com
  - Amazon.com
  - Reddit.com
  - Netflix.com
  - Creative Commons snippets are in RDFa

COMP318

Ontologies and Semantic Web



End of RDF - Part 6

Dr Valentina Tamma

V.Tamma@liverpool.ac.uk