Semantic Web Engineering

Gerald Reif
reif@ifi.unizh.ch

Fr. 10:15-12:00, Room 2.A.10

University of Zurich
Department of Informatics

s.e.a.l.
software evolution & architecture lab

Semantic Web

"The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation."

Tim Berners-Lee, James Hendler, Ora Lassila,
The Semantic Web, Scientific American, May 2001

- The well defined meaning is expressed by RDF meta-data.

➡We need annotated Web pages!

Semantic Annotation

- Aims to transform documents into machine-processable artifacts by augmenting them with meta-data that describes their meaning.
- Documents can be: Web pages, pictures, audio files, movies, etc.
- Writing OWL in RDF/XML syntax and typing URIs is error-prone.

➡ Tool support is needed

Association between HTML and RDF

- Once the RDF meta-data is generated the document and its meta-data have to be associated.
- Several techniques exist to associate the HTML Web page and its RDF meta-data description.
  - Embed RDF in HTML
  - Reference external RDF from HTML
  - RDFa for embedding RDF in XHTML (W3C working draft)
- There is no standard association technique.
- Proposed techniques are discussed in:
  - http://infomesh.net/2002/rdfinhtml/

Embedding RDF in HTML

- RDF/XML is directly embedded in the HTML page.
- We have to take care, that the annotations are not displayed by the Web browser.
- One possibility is to add RDF in the `<script>` element in the `<head>`.
- The mime-type `application/rdf+xml` is defined in RFC3870.

```
<head>
  <title>My Document</title>
  <script type="application/rdf+xml">
    <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://www.w3.org/"
                       dc:title="W3C Homepage"/>
    </rdf:RDF>
  </script>
</head>
```

HTML reference to external RDF

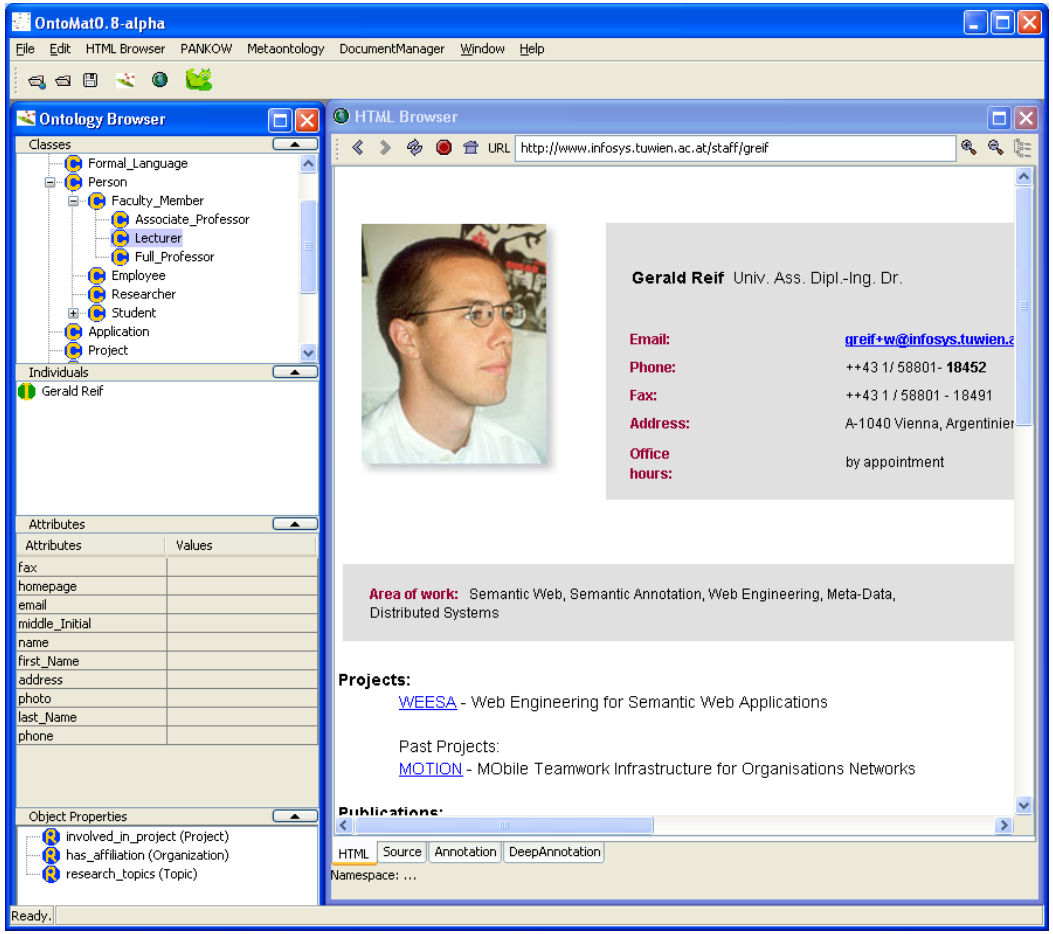
- The HTML page linkes to an external document with the RDF desription of the content.
- Recommended approach by the W3C: the use of the HTML `<link>` element:

```
<link rel="meta" type="application/rdf+xml"
                 href="Meta-Data-for-Web-Page.rdf"/>
```

- Others use a simpe HTML <a> link and the RDF logo for the link.

CREAM/OntoMat

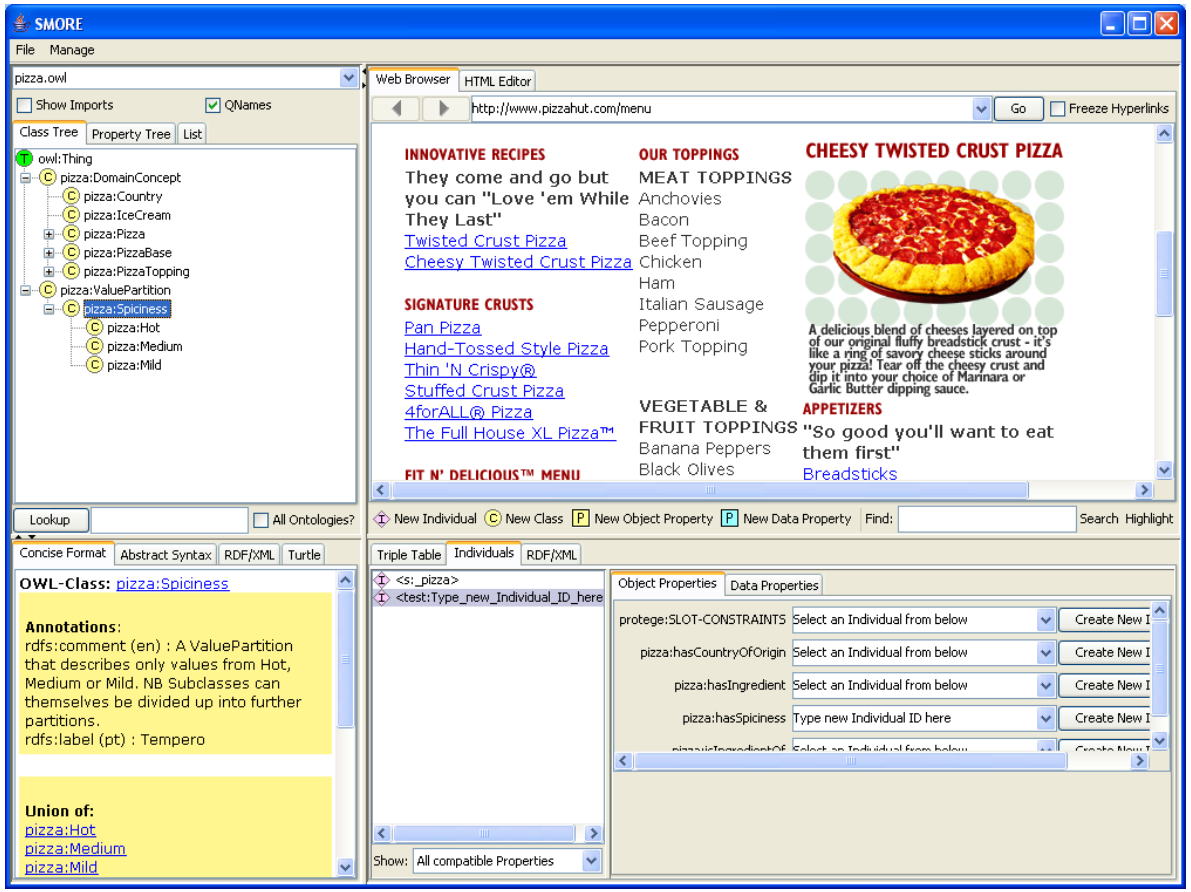
- A Semantic annotation framework that offers several annotation methods:
  - manual annotation
  - authoring of annotated documents
  - semiautomatic annotation
  - annotation of dynamic pages



http://annotation.semanticweb.org/ontomat

SMORE

- Tool to annotate HTML Web pages
- Uses OWL ontologies
- Built-In HTML editor
- Advanced ontology browser
- Developed in the Mindswap project



http://www.mindswap.org/2005/SMORE/

RDFa - RDF in XHTML

- RDF meta-data integration into XHTML Web pages.
- RDFa W3C Note
  - http://www.w3.org/TR/xhtml-rdfa-primer/
- RDF/A Syntax
  - A collection of attributes for layering RDF on XML languages
  - W3C Recommendation 14 October 2008
  - http://www.w3.org/TR/2008/REC-rdfa-syntax-20081014/
- Use RDFa annotations Examples
  - http://www.w3.org/2001/sw/BestPractices/HTML/rdfa-bookmarklet/

RDFa

- RDFa is a syntax that accomplishes metadata expression using a set of elements and attributes that embed RDF in XHTML.
- An important goal of RDFa is to achieve this RDF embedding without repeating existing XHTML content when that content is the metadata.
- An XHTML document marked up with RDFa constructs is a valid XHTML Document.

Attributes used in RDFa

- `@about`: Defines the subject of the following triples. The subject is used for all nested elements. If no `@about` is given, the subject is a blank node.
- `@property`: Defines the predicate of the triple with a literal object.
- `@content`: Defines a literal object.
- `@resource`: Defines a resource object.
- `@typeof`: Defines the rdf:type of the subject.
- `@rel`: Defines the predicate of a triple with a resource object. Object is defined using `@resource` or content from `@href` or `@src` attributes is reused.
- `@rev`: Defines the reverse property that uses the `@resource`, `@href` or `@src` reference as subject and the resource defined in the `@about` attribute as object.
- `@datatype`: Defined the data type of a literal.

RDFa: Blank Node Suject

Add namespace for annotation

```
<html xmlns:cal="http://www.w3.org/2002/12/cal/ical#">
    <head><title>Jo's Blog</title></head>
    <body>
...
    <p typeof="cal:Vevent">
        I'm giving
        <span property="cal:summary">
            a talk at the XTech Conference about web widgets
        </span>,
        on
        <span property="cal:dtstart" content="20060508T1000-0500">
            May 8th at 10am
        </span>.
    </p>
...
    </body>
</html>
```

Define the Resource to describe (Subject)

Define Property and Literal taken from the text

Define Property and Literal taken not available in the text

Defined RDF Graph

```
_:p0 rdf:type cal:Vevent;
     cal:summary "a talk at the XTech Conference
                  about web widgets";
     cal:dtstart "20060508T1000-0500";
```

RDFa: Resource References

```
<p class="contactinfo" about="http://example.org/staff/jo">
    My name is
    <span property="contact:fn">
        Jo Smith
    </span>.
    I'm a
    <span property="contact:title">
        distinguished web engineer
    </span>
    at
    <a rel="contact:org" href="http://example.org">
        Example.org
    </a>.
</p>
```

Subject resource ID

Predicate definition when object is a resource

Reuse of the HTML href to define the object resource

RDFa Example 1/2

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML+RDFa 1.0//EN"
    "http://www.w3.org/MarkUp/DTD/xhtml-rdfa-1.dtd">
<html xmlns="http://www.w3.org/1999/xhtml"
    xmlns:foaf="http://xmlns.com/foaf/0.1/"
    xmlns:dc="http://purl.org/dc/elements/1.1/”
    base="http://example.org/john-d/" xml:lang="en">
  <head>
    <title>John's Home Page</title>
    <span about="" property="dc:creator" content="Jonathan Doe" />
  </head>
  <body about="">
    <h1>John's Home Page</h1>
    <p>My name is <span property="foaf:nick">John D</span> and I like
      <a href="http://www.neubauten.org/" rel="foaf:interest"
        xml:lang="de">Einstürzende Neubauten</a>.</p>
    <p>My <span rel="foaf:interest" resource="urn:ISBN:0752820907">favorite
      book</span> is the inspiring <span about="urn:ISBN:0752820907"><cite
        property="dc:title">Weaving the Web</cite> by
      <span property="dc:author">Tim Berners-Lee</span></span>.</p>
  </body>
</html>
```

RDFa Example 2/2

RDF/XML included as RDFa in XHTML on previous slide:

```
<?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/"
    xmlns:dc="http://purl.org/dc/elements/1.1/">
  <rdf:Description rdf:about="http://example.org/john-d/">
    <dc:creator xml:lang="en">Jonathan Doe</dc:creator>
    <foaf:nick xml:lang="en">John D</foaf:nick>
    <foaf:interest rdf:resource="http://www.neubauten.org/"/>
    <foaf:interest>
      <rdf:Description rdf:about="urn:ISBN:0752820907">
        <dc:author xml:lang="en">Tim Berners-Lee</dc:author>
        <dc:title xml:lang="en">Weaving the Web</dc:title>
      </rdf:Description>
    </foaf:interest>
  </rdf:Description>
</rdf:RDF>
```

Defined RDF Graph

```
<http://example.org/staff/jo>
        contact:fn "Jo Smith”;
        contact:title "distinguished web engineer";
        contact:org <http://example.org>.
```

RDFa Examples

- RDFa example pages and tools can be found at
  - http://www.w3.org/2001/sw/BestPractices/HTML/rdfa-bookmarklet/
- RDFa Highlight for highlighting RDFa markup in XHTML pages.

Microformats 1/2

- Microformats are markup that allow expression of semantics in an HTML web page.
- Existing HTML standards allow for semantics to be embedded and encoded within them.
- This is done using specific HTML attributes:
  - class, rel, rev
- Not a W3C recommendation
- Example:

```
<span class="vcard">
     <span class="fn">Joe Doe</span>
     <span class="org">The Example Company</span>
     <span class="tel">604-555-1234</span>
     <a class="url" href="http://example.com/">http://example.com/</a>
 </span>
```

Microformats 2/2

- Several microformats have been developed to enable semantic markup of particular types of information.
  - hCalendar - for events
  - hCard - for contact information
  - hReview - for reviews
  - hResume - for resumes or CVs
  - rel-directory - for distributed directory creation and inclusion
  - rel-nofollow, an attempt to discourage 3rd party content spam (e.g. Spam in blogs).
  - rel-tag - for decentralized tagging (Folksonomy)
  - xFolk - for tagged links
  - XFN - for social relationships
  - XOXO - for lists and outlines

URIs as URLs – Content Negotiation to retrieve the correct data

- The same URI ist used for the HTML page and its RDF description
- HTTP content negotiation is used to specify the wanted content type.
  - `accept` in the HTTP request header set to `application/rdf+xml` for RDF metadata
- Example URI: http://xmlns.com/foaf/0.1/

```
$ telnet xmlns.com 80
Trying 66.33.211.178...
Connected to xmlns.com.
Escape character is '^]'.
GET /foaf/0.1/ HTTP/1.1
host: xmlns.com
accept: application/rdf+xml
```

HTML page is delivered when set to `text/html` or line is obmitted at all (server default response type)

```
HTTP/1.1 200 OK
Date: Fri, 05 Dec 2008 07:48:42 GMT
Server: Apache/2.0.61 (Unix) PHP/4.4.7 mod_ssl/2.0.61 OpenSSL/0.9.7e mod_fastcgi/2.4.2 Phusion_Passenger/2.0.2 DAV/2
     SVN/1.4.2
Vary: Accept
Last-Modified: Fri, 15 Aug 2008 11:58:19 GMT
ETag: "7b04b09-9c51-56664cc0"
Accept-Ranges: bytes
Content-Length: 40017
Content-Type: application/rdf+xml

<!-- This is the FOAF formal vocabulary description, expressed using W3C RDFS and OWL markup. -->
```