Kap. 10:
Schema.org and Semantic Markup

Eirik Folkestad

What is schema.org?

- RDF provides a model to represent structured data in a way that a computer can mechanically operate on it.
- Ontologies provides a set of shared terms to create RDF triples → RDF triples have no clear definition without.
- RDF data model and ontology → Semantic Web

Joint effort by Google, Bing and Yahoo!

Schema.org is an attemt to create one general ontology/vocabulary to make the benefits of the Semantic Web appreciated and understood by the general public.

Goal: Search engines understand page content → better search performance

Schema vocabulary

- Tries to cover all common aspects of our daily life -> Many categories and classes
- Supports several syntaxes
- The vocabulary of schema.org grows by absorbing other ontologies

Class

```
<owl:Class rdf:about="http://schema.org/Event">
        <rdfs:label xml:lang="en">Event</rdfs:label>
    <rdfs:subClassOf rdf:resource="http://schema.org/Thing"/>
    <rdfs:comment xml:lang="en">
        An event happening at a certain time at a certain location.
    </rdfs:comment>
</owl:Class>
```

Property of a Class

```
<owl:ObjectProperty rdf:about="http://schema.org/location">
        <rdfs:label xml:lang="en">location</rdfs:label>
        <rdfs:comment xml:lang="en">The location of the event or organization.</rdfs:comment>
        <rdfs:range>
            <owl:Class>
                <owl:unionOf rdf:parseType="Collection">
                    <rdf:Description rdf:about="http://schema.org/Place"/>
                    <rdf:Description rdf:about="http://schema.org/PostalAddress"/>
                </owl:unionOf>
            </owl:Class>
        </rdfs:range>
```

Property of a Class

```
        <rdfs:domain>
                <owl:Class>
                        <owl:unionOf rdf:parseType="Collection">
                                <rdf:Description rdf:about="http://schema.org/Event"/>
                                <rdf:Description rdf:about="http://schema.org/Organization"/>
                        </owl:unionOf>
                </owl:Class>
        </rdfs:domain>
</owl:ObjectProperty>
```

Content markup

Schema.org only provides the terms (types and properties) so we need a language to do the actual markup.

- Microformats
- Microdata
- RDFa / RDFa 1.1 Lite (Minimal subset of RDFa)

RDFa 1.1 Lite

Five attributes: vocab, typeof, property, resource and prefix

```
<div vocab="http://schema.org/" resource="http://www.liyangyu.com/foaf.rdf#liyang" typeof="Person">

// sets vocabulary to schema.org and also sets RDFa Lite as the markup language and class of div to Person. Resource is needed if the resource is explicitly needed

        <p>My name is <span property="name">Liyang Yu</span>, I live in // tags the content of the span as a property of Person

        <span rel="address"> // use rel if property is an instance of a class

                <span typeof="PostalAddress"> // define the instance that address takes as value

                        <span property="addressLocality">Atlanta</span>,

                        <span property="addressRegion">Georgia</span>.

                </span>

        </span>

        </p>

</div>
```

Other onthologies

```
<div vocab="http://schema.org/" prefix="foaf:http://xmlns.com/foaf/0.1/"
    … // Can use prefix to access other onthologies
    <span rel="foaf:knows">
        <span typeof="foaf:Person">
            <span property="foaf:name">Carmen Hetrea</span>
```

What format to use for schema.org?

- Microdata or RDFa? (Microformat: cannot make your own microformat if you want to mark up some kind of data that's not supported by the official specs)
- Google supports microdata → simpler than RDFa
- Google also supports RDFa Lite
- Likely to coexist for a long time, but RDFa has been seen growing steadily among the three and RDFa and RDFa Lite are both W3C standards

Issues

- Create structured data that is useful
  - Important to use right “return” type for a class. Even though the type is wrong any type can be allowed → so the required type is up to the developer to maintain.
  - The amount of markup is up to the developer → incomplete instances?
  - Invisible vs. visible markup → invisible markup have advantages but this is forbidden due to a potential risk of spam. There are some exceptions to this rule i.e. geographical locations
  - Time and date should be marked with a content attribute to avoid ambiguity i.e. 01/11/2013

Validating the markup

Validation tool:

https://www.w3.org/2012/pyRdfa/

http://www.imdb.com/title/tt3062096/?ref_=nv_sr_1

To validate whether the markup is correctly done, we need to inspect the generated file to see if it has all the triples that we expect to see.

How is the markup information being used?

- Rich snippets

- LMRI project

Rich snippets

Summarization and description of each individual result page in such a way that it gives the user a clear indication of how relevant this current page is to the original search. → Page snippet

A **rich page snippet** takes advantage of the semantic web technology to make a better page snippet. Many sites/companies using rich snippets have seen an increase by 20%-30% in click-through rates.

Google has its own validation tool for rich snippets.

Google supports rich snippets of these types (by the time the book was written):

- reviews
- people
- products
- business and organizations
- recipes
- events
- music

LRMI - The Learning Metadata Resource Initiative

While the Internet makes searching for learning resources easy, the sheer vastness of material available makes it extremely challenging to find just the right learning resources to address specific student needs.

The LMRI project is created to address this problem.

The LMRI specification is adopted into the vocabulary of schema.org.

ISLE - Open Education Resources

LMRI specification:

- subject/area
- age range
- standards alignments
- media type
- interactive type (such as active, mixed)
- time required
- usage type (such as assignment or group work)
- resource type (such as handout or presentation)
- copyright/licensing information