Global / Integrated Access and Ontologies

Information, Knowledge, Libraries & the 'Semantic Web': Concepts, Challenges & a Proposal

Dr. Stefan Gradmann
Regionales Rechenzentrum der Universität Hamburg
stefan.gradmann@rrz.uni-hamburg.de
www.rrz.uni-hamburg.de/RRZ/S.Gradmann
Overview

- The WWW before the 'semantic turn'
- What is the 'Semantic Web'? Does it actually exist?
- Where do 'ontologies' fit in the SemWeb framework?
- Some examples ...
- Conceptional confusions: data, information, knowledge
- Philosophical pitfalls: positivist vs. hermeneutical perspectives of truth
- What to expect from ontologies - and what not?
- The potential impact on 'libraries'
- A concluding proposal – frbr:rdfs – and another look on libraries
The WWW before the 'semantic turn'

- A curious blend of metaphors: a 'network' with 'sites', 'pages' and 'documents'
- A data accumulation and transport machine
- A mere carrier medium
- 'Brute force' operations on the accumulative and on the selective side resulting in
- A fragile information economy and (at times)
- Coming close to 'white noise'
The 'SemWeb' according to Artur, Crofts & Le Boeuf

Web + Knowledge Representation
= Semantic Web
or
knowledge structures for the sake of machines
The 'SemWeb' according to Berners Lee

<table>
<thead>
<tr>
<th>Rules</th>
<th>Data</th>
<th>Self-Desc. document</th>
<th>Digital signature</th>
<th>Lexicon, Syntax</th>
<th>Carrier</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Content: Semantics</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trust</td>
<td>Proof</td>
<td>Logic</td>
<td>Ontology vocabulary</td>
<td>RDF + rdfschema</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>XML + NS + xmlschema</td>
<td></td>
</tr>
<tr>
<td>Unicode</td>
<td>URI</td>
<td>HTTP, URL, HTML</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Rules**
- **Data**
- **Self-Desc. document**
- **Digital signature**
- **Lexicon, Syntax**
- **Carrier**

**Technologies:**
- Unicode
- URI
- XML + NS + xmlschema
- RDF + rdfschema
- HTTP, URL, HTML

The 'SemWeb' according to Berners Lee.
What is an 'ontology'? 

- An **ontology** is a formal explicit description of **concepts** in a domain of discourse (**classes**), **properties** of each concept describing various **features** and **attributes** of the concept. An **ontology** together with a set of individual **instances** of classes constitutes a **knowledge base**.

- **Basic elements are S -> P -> O triplets:**
  
  'speaker' -> 'HasName' -> 'Gradmann'
What does an ontology it look like?

- Basic elements are S -> P -> O triplets

Example* (small fragment of a knowledge base, visualized):

*from Noy/McGuinness 2001
Semantics, WWW and ontologies: two widespread confusions (I)

Depending on what level we are operating on we conceive networked bits & bytes as

- Data or as
- Information or as
- Knowledge

- and all too often implicitly confound these conceptual layers
Semantics, WWW and ontologies: two widespread confusions (II)

Carrier <-> content / signifier <-> significate

When mentally organizing 'content' what are we actually referring to:

- 'things',
- pointers to 'things',
- meta-'things',
- 'signs'

- here again confusion is widespread!
Ontologies and assumptions about 'truth': philosophical pitfalls

Benel, Aurélien [et al., 2001]* distinguish three metaphysics of truth:

- **Positivist** (a priori, consensual, assuming 'obvious' things)
  vs.

- **Conventionalist** and **Hermeneutical** paradigms of 'truth'.

=> 'Ontologies' only make sense in the first paradigm!

=> Are the others excluded from WWW knowledge management?

* Truth in the Digital Library. From Ontological to Hermeneutical Systems (LNCS 2163)
What use can be made of ontologies?

- Model terms that are equivalents of those sections of our 'world' we feel able to organize in taxonomic models of knowledge organization.

- In those cases that does not primarily require 'interpretation'.

- For 'evident' domains e. g.
  - Places
  - Time periods
  - Non-cultural artefacts such as cars and aeroplanes ...

- Everything that can be ... classified!

- But within a specific archicietural framework!

- And as a basis for automated operations on content!!
And the potential impact on 'libraries'

Very much depending on what mission 'Digital Libraries' will ultimately have and the way they will choose to use the WWW:

- **Basically self contained content stores using proprietary metadata standards** and the WWW for data transfer (focus on carrier)?
  => will not need to be involved with 'ontologies' and SemWeb at all (but may face serious other problems because of this choice!)

- **Content stores integrated in WWW content architectures with WWW-transparent metadata standards** (focus on syntax)?
  => limited, yet systematic impact of SemWeb technology and ontologies, as exemplified in the concluding section

- **Content store and enabler for content based operations** (focus on semantics)?
  => SemWeb technology would become core business of such 'libraries' (but this remains a very utopist and unlikely scenario ...)

12
Figure 3.1: Group 1 Entities and Primary Relationships

WORK

is realized through

EXPRESSION

is embodied in

MANIFESTATION

is exemplified by

ITEM
An Excursion to FRBR (II)
... and a proposal, frbr:rdfs

A potential future way of defining and implementing 'cataloguing rules' might look like this:
Some benefits of frbr:rdfs

- Make librarian metadata resources WWW-transparent which are today part of the 'hidden web'
- Do so without drowning the WWW with heavily redundant cataloguing information
- Potential of inference based models for generating cataloguing and indexing information
- Make the classification and thesaurus building work of generations of librarians available as a resource for ontology building
- Render obsolete the traditional distinction between descriptive and subject metadata in librarian workflows
... and the place of ontologies

And in such a setting the place of ontologies would be well defined!

One example of where Ontologies might fit in:

```xml
<?xml version='1.0' encoding='UTF-8'?>
<!DOCTYPE rdf:RDF [ 
  <!ENTITY a 'file:/F:/apps/Kaon/ontologies/frbr#'> 
[...] xmlns:rdfs="&rdfs;">
<rdfs:Class rdf:ID="place">
  <rdfs:label xml:lang="en">place</rdfs:label>
  <rdfs:subClassOf rdf:resource="#work"/>
  <rdfs:subClassOf rdf:resource="#Corporate-Body"/>
</rdfs:Class>
[...]
</rdf:RDF>
```

Places ontology
Ontologies in 'librarian' contexts

- **Yes**, useful: if embedded in a WWW transparent information architecture

- **No**, useless effort: if integrated in the library automation paradigm as still basically valid today

- **Yes**, can save a lot of human resources for intellectually demanding tasks: if the limitations are well understood

- **Not**: if regarded as panacea for all problems of structuring, interoperability and reuse of librarian information models

- **Not** if viewed as continuing business (classification) using other means, and thus not as part of 'Digital Library' settings!

- **Not** if this means once again stepping in the traps of AI!

... and the majority of negative statements can lead to two very different conclusions!

Thank you for your patience and attention!