Developing an Ontology for Academic Disciplines

Martin Stricker
Hermann von Helmholtz-Zentrum für Kulturtechnik
Humboldt-Universität zu Berlin
15 October 2008
Hermann von Helmholtz-Zentrum für Kulturtechnik

- An interdisciplinary research center at Humboldt University, Berlin

- „Kulturtechnik“ – the study of scientific and cultural shifts in the context of technical innovations

- Focal point (among other things) is communication in and of science and especially the role of „objects“ and scientific collections therein
Serie: Exilbriz, Rudolf Virchow

Fotografie: Der Fußkünstler Uthman mit Geige

Fotografie: Präparierübungen

Foto: Hans Virchow beim Präparieren

Objekt des Monats 3/2008

Wenkärzte in einer hessischen Mundart
**Objektgruppen Auswahl**

Im Index „Objektgruppen“ ermöglichen innerhalb der Domänen *res naturalia* und *res artificialia* Beschreibungsaspekte den Zugang zu relevanten Gruppen von Objekten aus unterschiedlichen Perspektiven.

### Domänen und Beschreibungaspekte

<table>
<thead>
<tr>
<th>Domäne</th>
<th>Beschreibungaspekte</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>res naturalia</em></td>
<td>Kunstgattung, Funktion/Form, Verwendungszweck, Lehrmaterial/Demonstrationsmaterial, Medium, Forschungsbereich, Material, Technik/Herstellungsort, Darstellungsthema, Entstehungszeit, Herkunftsregion, Zusatzinformation</td>
</tr>
<tr>
<td><em>res artificialia</em></td>
<td>Kunstgattung, Angewandte Kunst, Architektur, Architekturzeichnungen/Architekturpläne, Bildende Kunst, Buchkunst, Fotografie, Grafik, Druckgrafik, Zeichnung, Zeichnungen, Installation, Malerei, Wandmalerei, Plastik/Skulptur, Kleinplastik, Reliefs, Textilkunst, Darstellende Kunst, Bühnenbilder, Requisiten, Design, Musik</td>
</tr>
</tbody>
</table>

### Auswahl

<table>
<thead>
<tr>
<th>Auswahl</th>
<th>Beschreibung</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>res artificialia</em></td>
<td>Angewandte Kunst, Kunstgattung</td>
</tr>
</tbody>
</table>

### 3 Sammlungen · Aufrufen
Objektgruppen Auswahl


Mikrobiologie (2)

Oberbegriff
BT Biologie

Unterbegriff
NT Phylogenie

Benutzt für
UF Microbiology
Limitations of common Knowledge Organisation Systems and implementations

- Knowledge engineering level: Limited expressivity

- Implementation level: Each vocabulary has its own structure on the application level (application code, encoding or database schema)

- Interoperability level: Sharing or synchronizing information or data requires for each instance dedicated application code, and very often dedicated user interfaces for manual mapping
Solution for the domain of Science and Humanities

- An "ontology" which respects the qualities of traditional knowledge organisation systems (poly-/hierarchies, synonyms), enhanced by additional (syntagmatical) relation types and logical inferencing potentials

- The model enables the description of disciplinary characteristics and interdisciplinary connections via properties/relations of typed significant constituents

- Ontology encoded in RDF-based ontology language OWL (interoperability for information exchange, chance of reusable application code)
Design characteristics

- Terminological (linguistic) approach (currently rather low degree of formalization, lightweight inferencing)

- Basic usage shall support intuitive browsing and navigation, serendipitous discovery, assisted search (including entry points for more complex rule-based operations and representations)

- Main principle: As simple and intuitive as possible, as complex as necessary
Usage scenarios

- Indexing, browsing, (assisted or faceted) searching of collections and databases which contain data on a supra- or multidisciplinary level
  - Bibliographic information (library catalogs), research databases, university information systems, outside representation ...

- Integration of legacy vocabularies possibly from heterogenous sources

- Wider integration within the Semantic Web / Linked Data structures
Project team

Lead
Dr. Cornelia Weber, General Manager of Helmholtz-Zentrum

Principal developers
Dipl.-Phil. Information Specialist Elena Semenova
Concept, modelling

Martin Stricker
Semantic technologies & standards, software development
Class properties: Discipline

Group of Disciplines

Method

Usage

Part

Explanandum

Sub-/Superdiscipline
(skos:broader, skos:narrower)

Object of Research

Problem

Object

Technology

Formation

Usage

Focus

Chair (Lehrstuhl)

Course of Studies

Scientist

Research Institution

University

Teaching

Focus

Formation

Usage

Focus

Formation

Usage

Focus
Specifications

- OWL 1 (recommendation), SKOS (latest working draft)
- Usage of class descriptions (local property restrictions) / property descriptions for consistency enforcing & checking
- Usage of SKOS:
  - Hierarchical relations
  - (Natural language) names or labels (skos:prefLabel, skos:altLabel, skos:hiddenLabel)
Challenge: Communication

- Interdisciplinary communication is hard (especially in a project about interdisciplinary communication)

- Terms and meaning:
  - Class, Instance, Individual
  - Concept, Resource, (Topic, Association)
  - Name, URI, IRI, Label
  - plus the finer points of <xml>, Namespaces, First Order Logic, UML graphs, Open-World Assumption, Unique-Name Assumption
Challenge: Contextual statements

- RDF only allows binary relations or properties (relations between two resources/individuals)

- How do you model this: "In narratology, temporal and spatial qualities are aspects in the study of narratives"

- We had to „reify“ relations/properties as auxiliary resources (class/individual)
Challenge: Vocabulary Mapping

- How to integrate other (legacy) vocabularies and map their terms to concepts of the ontology?
  - Schlagwortnormdatei (SWD), DDC Deutsch (via CrissCross SWD-DDC Mapping), Regensburger Verbundklassifikation (RVK), Personennamendatei (PND), Gemeinsame Körperschaftsdatei (GKD)

- Missing namespaces and RDF representation

- Legal problems (copyright)

- For the actual mapping, we created auxiliary mapping classes in order to be able to annotate mapping relations (SKOS Mapping Properties can be auto-generated via rules/reasoner)
Other challenges

- Tools and software for ontology editing: Reliability, performance, permanent „beta“, ...
- Semantic (re-)factoring/ further formalization, inferencing
- Internationalization
- Modelling and describing timelines and historical processes
- Implementation strategies for modular remote access: Client/server-architecture, interfaces, API, protocols, ...
Further development: Core Ontology

- Refine and refactor model
  - Explore formalization options, logical rule-based reasoning

- Build a core vocabulary for a description of the German Science and Humanities disciplines (stable/long-term)
  - This in itself could trigger an application, a kind of modern interactive lexicon, tapping into other resources (Linked Data, Wikipedia/dbpedia)
  - The core vocabulary can be used as reference/background for applications which index specialised data or metadata for various purposes
Further development: Projects and implementations

- Cooperate with partners who have a need or see an option to develop, use or integrate our model & vocabulary

- Consulting or development:
  - Model extension
  - Vocabulary development
  - Integration of legacy data
  - Software development (especially modular or service-based approaches)