DERI INNSBRUCK

Leopold Franzens
Universität Innsbruck

GenTax: A Generic Methodology for Deriving OWL and RDF-S Ontologies from Hierarchical Classifications, Thesauri, and Inconsistent Taxonomies

Martin HEPP
DERI Innsbruck – University of Innsbruck
Jos de Bruijn
Faculty of Computer Science – Free University of Bolzano

making semantics **real.**

Goal: Derive OWL and RDF-S ontologies from any hierachical classification

- Being able to derive consistent RDF-S,OWL, and WSML ontologies from hierarchical classifications
- High degree of automation, i.e., without the need for manual analysis of conceptual elements
- Ability to transform SKOS vocabularies into RDF-S, OWL, or WSML

Motivation: Hierarchical classifications as a major resource

- Hierarchical classification systems are a major resource for structuring information
- Well established means in information management
- However, reuse for building ontologies not trivial
  - Fuzzy notion of class membership
  - Context-dependent semantics

Assets for the Semantic Web: A Wealth of Consensual Concepts

- **UNSPSC,** http://www.unspsc.org
  - 20,700 classes, 55 top-level categories
- **eCl@ss,** http://www.eclass.de
  - 25,000 classes, 25 top-level categories
- **eOTD,** **http://**www.eotd.org
  - 59,000 classes, 79 top-level categories

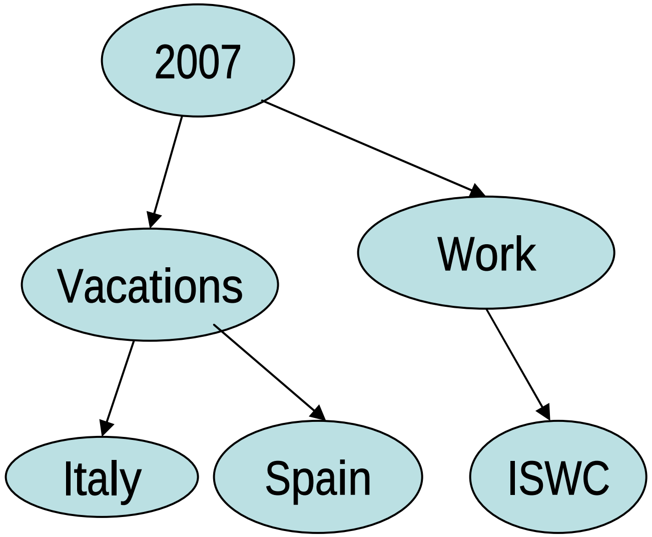
- DMOZ
- Wikipedia Categories

etc.

Definition: Hierarchical classifications

We view a **hierarchical categorization schema** as

- **a directed graph**
- where **nodes represent categories** and
- **edges represents the "narrower term"** or "has subcategory" relation.
- Depending on the **context**, a set is related to each category.
- This set represents the **items associated with the category in a particular context.**



Observations

- Classifications do not require a context independent definition of their intended meaning
  - We can use the same classification in multiple distinct contexts with very different meanings, as long as we keep the contexts apart
    - Invoices vs. Documents vs. Tasks
- When we use the labels as definitions for sets, we interprete them over the context of usage
- The meaning of a label and the meaning of the edges are mutually dependent
- We may face several anomalies

Anomalies

- **Local labels**
  - Pictures->Italy->Summer 2006
  - Pictures-> Pictures.Italy -> Pictures.Italy.Summer_2006
- **Hierarchy:** Depending on the context over which we interprete a label, the original set of arcs
  - **may** or
  - **may not**

  constitute a valid subsumption hierarchy

The Semantics of Category Labels

1

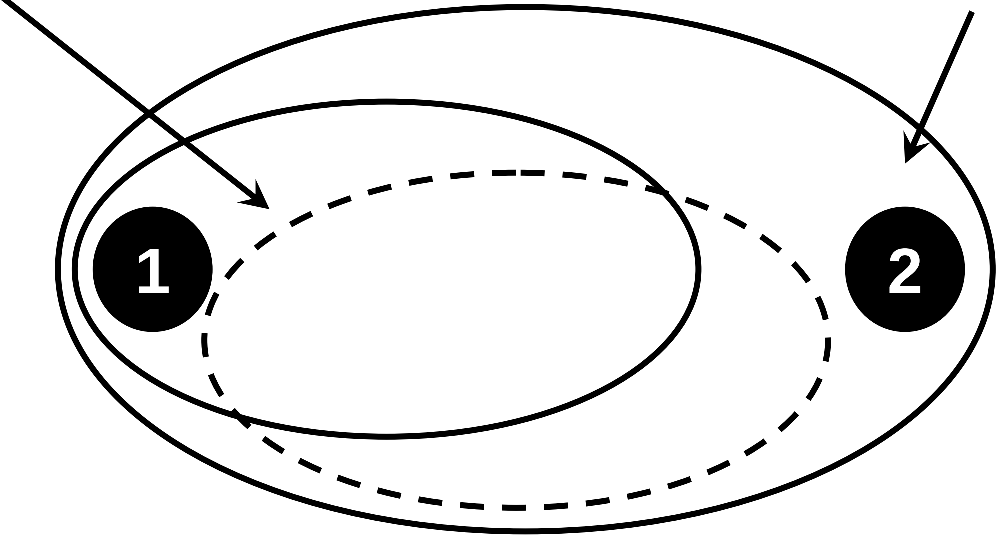
Concept in some context

(Example: “*Pictures of* Italy”)

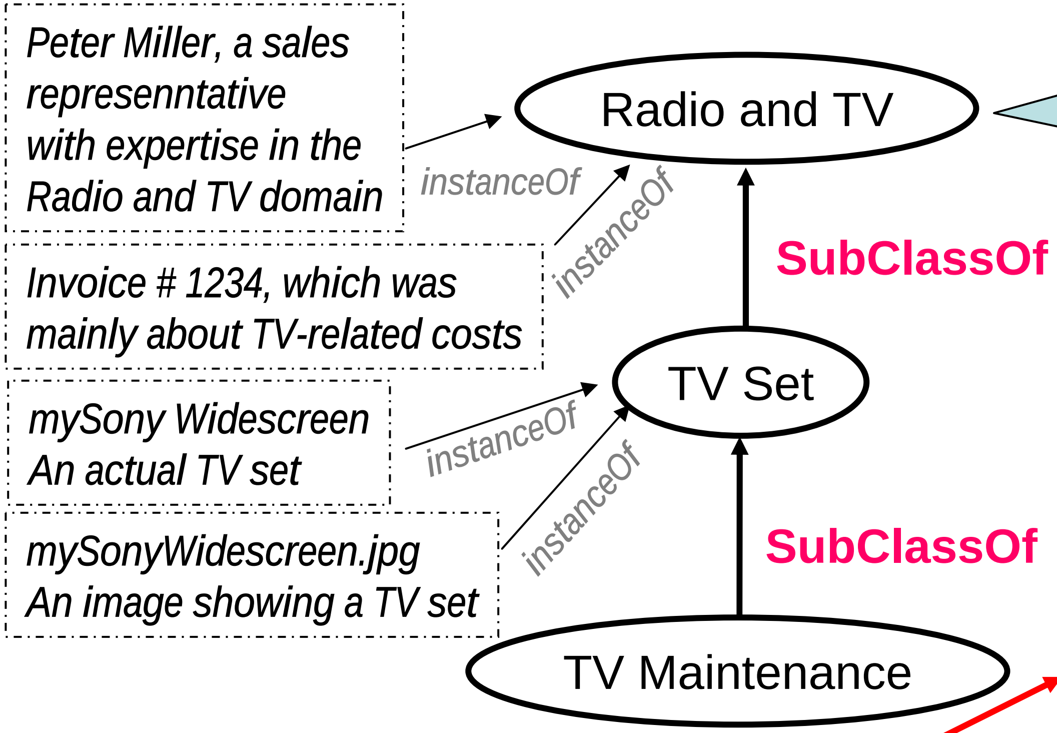
2

Category Concept

(Example: “**Anything** that can **in any reasonable context** be subsumed under the label “Pictures.Italy”)



Naïve Approach: One ontology class per each label (leads to ontologies limited in use)



Peter Miller, a sales representntative with expertise in the Radio and TV domain

Invoice # 1234, which was mainly about TV-related costs

mySony Widescreen An actual TV set

mySonyWidescreen.jpg An image showing a TV set

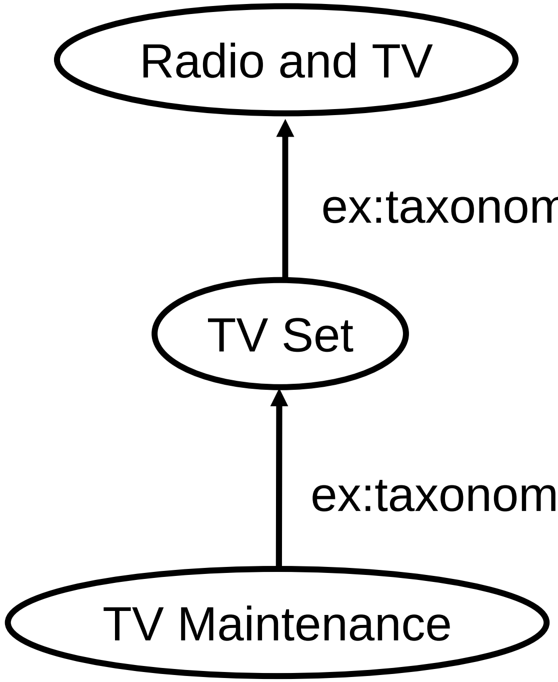
Example: “Anything that can in any reasonable context be subsumed under the label “Radio and TV”

Example for two contexts that differ in whether the original edges would constitute a valid subsumption hierarchy!

Examples for Contexts:

- Products
- User‘s Manuals
- Employees with Sales Expertise

Solution 1:One class for each node in the target context (but impossible in OWL)



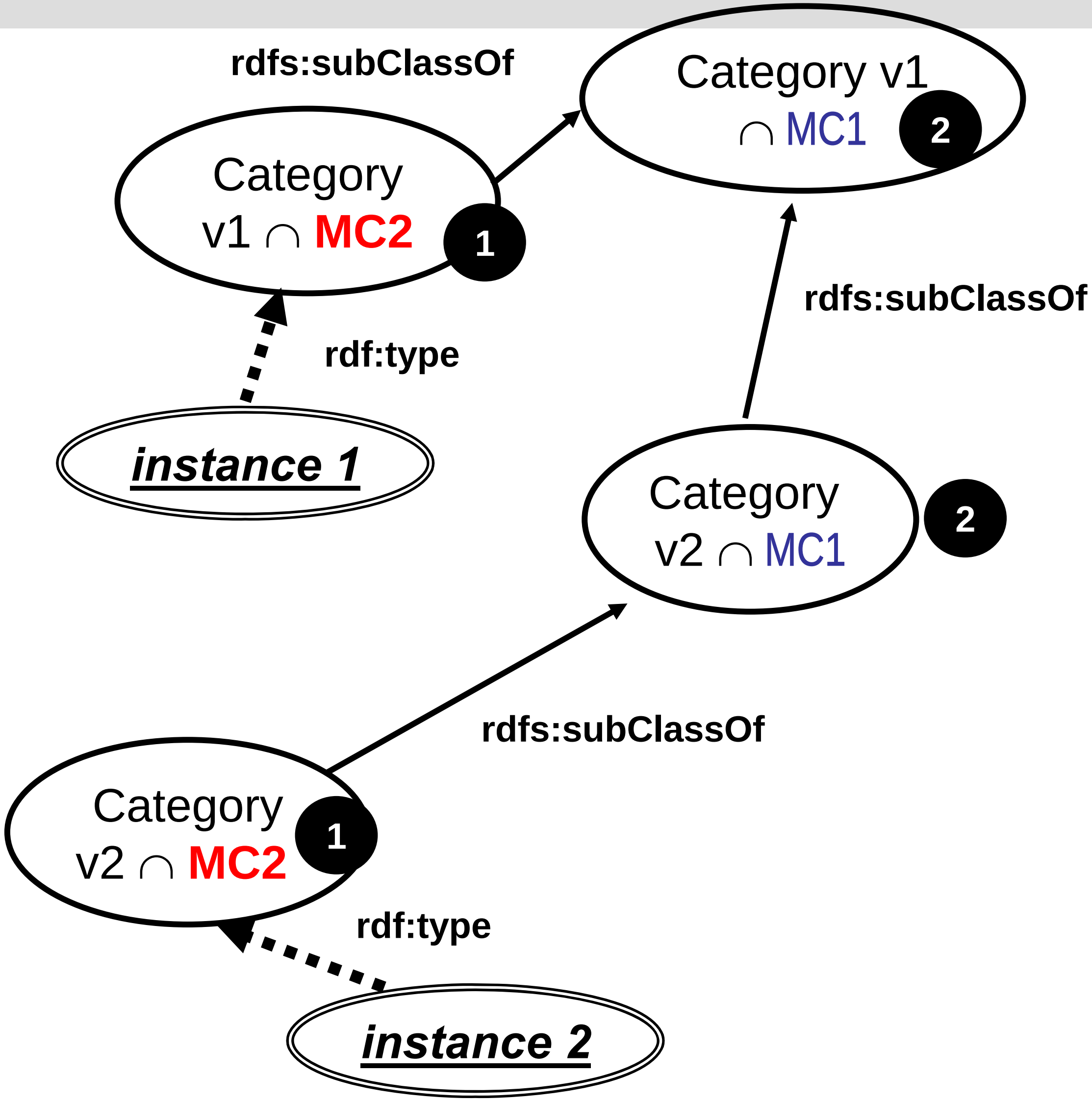
„Meant as an actual product described by this label“

Still, you have to select one target context!

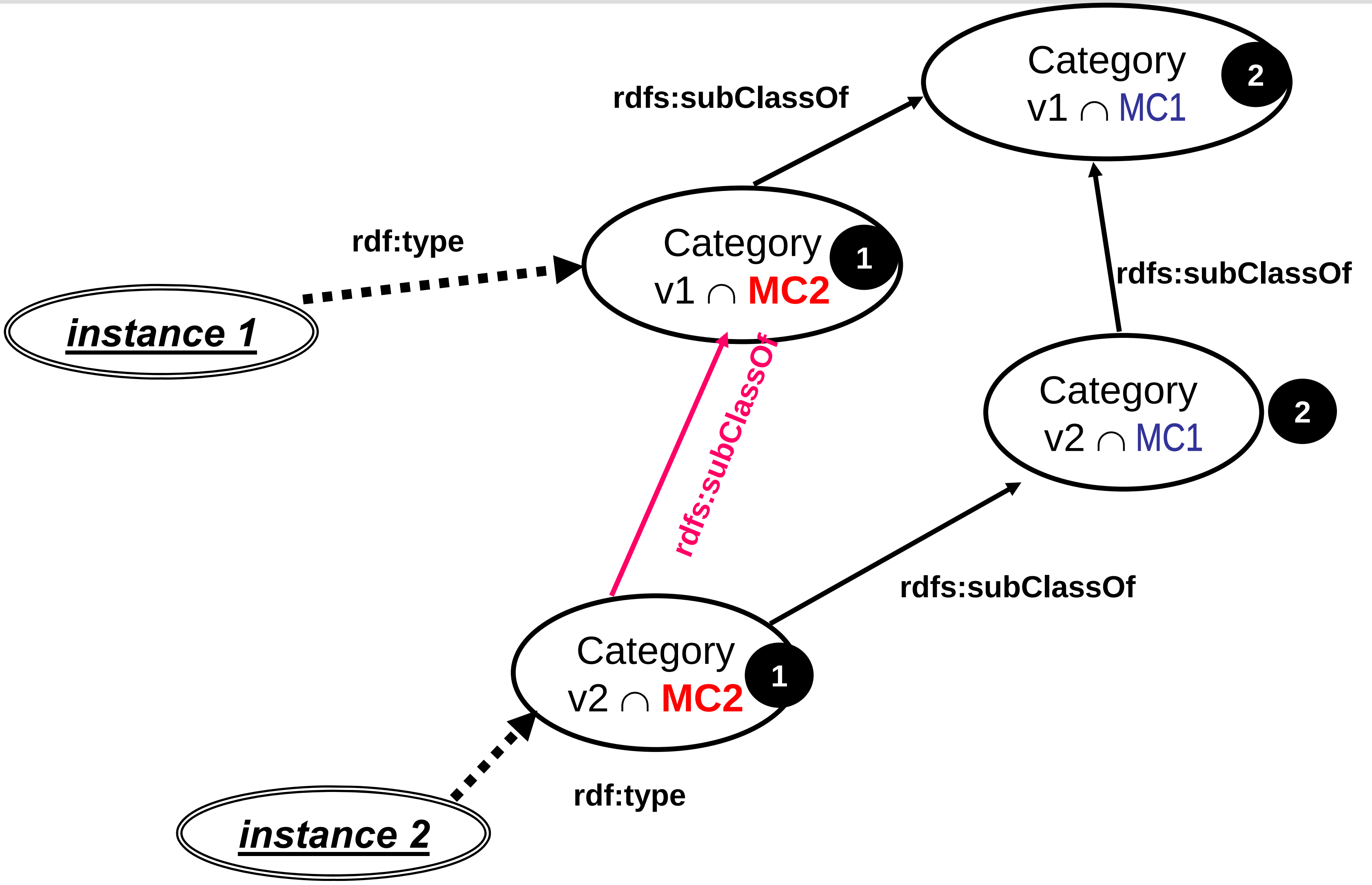
FORALL
A ex:taxonomySubClassOf B AND
B ex:taxonomySubClassOf C
→ A ex:taxonomySubClassOf C

The GenTax Idea

1. Define two Master Concepts:
   a) MC1 – any item for which the orginal hierarchy was intended.
   b) MC2 – the set of all entities in the target context
2. Check whether the orginal hierarchy is a valid subsumption hierarchy if you interprete the categories as label ∩ MC1
3. Check whether each label ∩ MC2 is a proper subclass of this label ∩ MC1
4. For steps 3 and 4, we take representative samples only!



Special (Lucky) Case



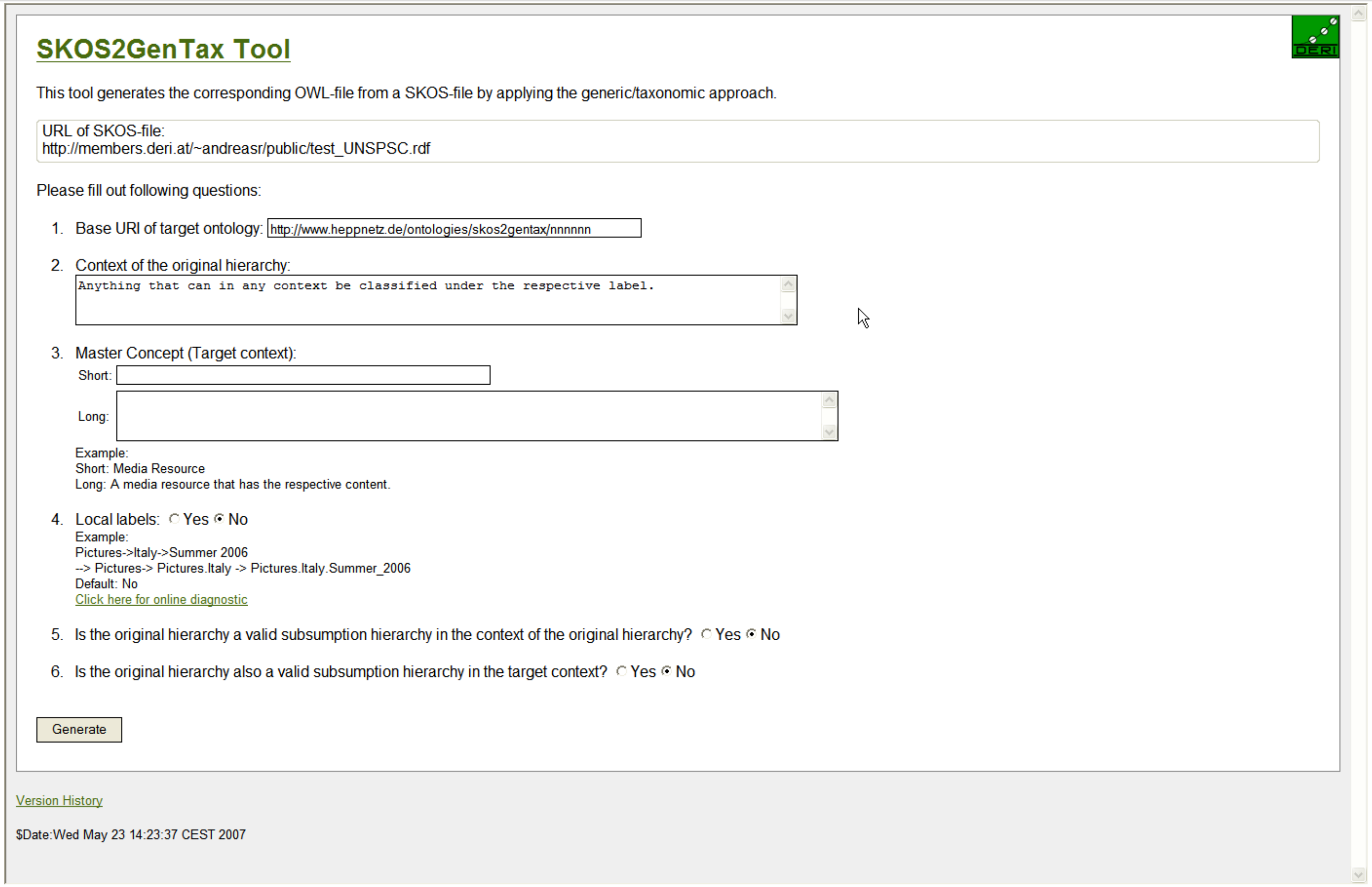
Nice Properties

- Automatic creation of lightweight ontologies possible that require only subClassOf as a modeling element
  - Resulting ontologies can be expressed in most popular ontology languages
- Original hierarchy can be preserved while still being able to design more specific ontology classes for each label
- Only a small sample necessary to decide upon proper conceptual modeling
  - No need to manually analyze each single element of large classifications

Statistical Diagnosis of Anomalies and Modeling Choices

- Idea
  - We draw a representative sample of the input classification
  - We ask a human to decide for this small sample whether for this element certain modeling choices are correct
    - e.g. whether, as categories for expenses, TV maintenance is a subclass of TV Set
    - same for local labels and other anomalies
  - We accept or reject that modeling choice for the full classification based on the sample
- Advantages
  - We have a solid statistical basis for the decision
  - We can chose a suitable degree of confidence depending on the target domain of the ontology
    - classifying Web documents vs. life sciences

Tooling: SKOS2GenTax Converter



SKOS2GenTax Tool

This tool generates the corresponding OWL-file from a SKOS-file by applying the generic/taxonomic approach.

URL of SKOS-file:
http://members.deri.at/~andreasr/public/test_UNSPSC.rdf

Please fill out following questions:

1. Base URI of target ontology: http://www.heppnetz.de/ontologies/skos2gentax/nnnnnn
2. Context of the original hierarchy:
   Anything that can in any context be classified under the respective label.
3. Master Concept (Target context):
   Short:
   Long:
   Example:
   Short: Media Resource
   Long: A media resource that has the respective content.
4. Local labels: ○ Yes ● No
   Example:
   Pictures->Italy->Summer 2006
   --> Pictures-> Pictures.Italy -> Pictures.Italy.Summer_2006
   Default: No
   Click here for online diagnostic
5. Is the original hierarchy a valid subsumption hierarchy in the context of the original hierarchy? ○ Yes ● No
6. Is the original hierarchy also a valid subsumption hierarchy in the target context? ○ Yes ● No

Version History

$Date:Wed May 23 14:23:37 CEST 2007

Will be online shortly at http://www.heppnetz.de/skos2gentax/

Full Paper

Martin Hepp, Jos de Bruijn: *GenTax: A Generic Methodology for Deriving OWL and RDF-S Ontologies from Hierarchical Classifications, Thesauri, and Inconsistent Taxonomies,* Proceedings of the 4th European Semantic Web Conference (ESWC 2007), June 3-7, Innsbruck, Austria, in: E. Fraconi, M. Kifer, and W. May (Eds.): ESWC 2007, LNCS 4519, Springer 2007, pp.129-144.

Thank you.

Martin HEPP
DERI Innsbruck – University of Innsbruck
Jos de Bruijn
Faculty of Computer Science – Free University of Bolzano