

A Survey and Categorization of Ontology-Matching Cases

Zharko Aleksovski

Philips Research
zharko.aleksovski@philips.com

Willem Robert van Hage

TNO Science & Industry
wrvhage@few.vu.nl

Antoine Isaac

Vrije Universiteit Amsterdam
aisaac@few.vu.nl

1. Research Questions

1. What are the different kinds of cases in which ontology matching has been deployed so far?
2. Can we observe common aspects, leading to a classification of these cases?
3. Are there matching techniques that have been observed to perform better on specific sets of cases?

3. Case Descriptions

For the full case descriptions see:
<http://www.few.vu.nl/~aisaac/iswc2007/cases>

Agricultural Thesauri

NAL Agricultural Thesaurus and AGROVOC
During OAEI 2006 the thesauri of the US Dept. of Agriculture (USDA), and the United Nations Food and Agriculture Organisation were linked by equivalence mappings, so as to allow integrated search over the literature databases of both institutes.

MACS

SWD, Rameau, LCSH
Three libraries' subject heading lists are aligned to provide multilingual subject-based access to their collections. The search engine uses mappings to propose users reformulations of their queries.

Renardus integrated gateway

10 vocabularies + DDC
Integrated access to collection gateways is provided by means of query reformulations and browsing along the DDC class hierarchy, which was aligned to the specific collections' vocabularies.

STITCH integrated Cultural Heritage browser

Iconclass, ARIA catalogue
A browser allows users to access two CH collections by one vocabulary or another, using the hierarchical structure of the vocabularies to define their search.

Vascoda

27+ vocabularies, LCSH, MeSH, AGROVOC, SWD, etc.
Vascoda is a german portal for scientific publications that offers meta-data search. The mappings are used to provide federated access to documents from over 23 libraries.

WebDewey terminology service

DDC, LCSH
This terminology service provides an access the the DDC classification scheme. This includes links to headings from LCSH so that users can use LCSH headings to better find and understand DDC classes.

Intensive Care

OLVG and AMC controlled vocabularies
Mapping in the form of pairs of corresponding OLVG and AMC hospital vocabularies is required in order to migrate the patient data from one system to the other.

5. Useful Techniques by Problem Type

We consider four different types of matching techniques that were used in the different cases:

- **Lexical:** based on lexical comparisons of labels and glosses
- **Structural:** using the structure of the ontologies
- **Background knowledge:** using additional external knowledge
- **Instance-based:** using classified instance data.

Based on the above cases, we observe that some techniques turn out to be more useful for some case categories than for others:

Category of use cases	Lexical	Structural	Background knowledge	Instance-based
Question answering	✓	✓		
Unified view over collections	✓	?	✓	?
Serendipity in browsing		✓		✓
Data migration	✓		✓	

2. Case Selection Criteria

We only selected cases that:

1. provide information on the actual alignment
2. report on which techniques can be used to solve the matching problem
3. clearly describe how the correspondences will be used in an application.

High-Performance Knowledge Base

70+ knowledge bases
Mapping is used for question answering to support intelligence experts in making strategic decisions

Unified Medical Language System

SNOMED Clinical Terms, Medical Subject Headings (MeSH), The Gene Ontology (GO), UWDA, NCBI Taxonomy, etc. (100 vocabularies)
Mapping between biomedical vocabularies from several organizations is used for several purposes, among which one is question answering.

Internet Directories

Google and Yahoo! directory
Mapping is needed for the purpose of browsing through interconnected directories from Internet portals.

Internet Music Schemes

CDNOW, MusicMoz, Artist Direct Network, All Music Guide, Artist Gigs, CD Baby and Yahoo LaunchCast hierarchies
Mapping is needed for the purpose of serendipity in mixed browsing of different music content schemes. When browsing through a scheme to find artists, releases and songs, users would like to access the content of corresponding or related classes in other providers' schemes.

4. Case Classification by Problem Type

Two important criteria to analyse an ontology mapping case are:

- **Precision vs. Recall:** for some cases assessing a mapping is time consuming, for others it is very time consuming to find a missing correspondence
- **Complexity of Representation:** some use cases only need basic semantic structures, others need rich ontologies with many different properties and use of logical axioms.

Based on the above cases, we observe that use cases with the same problem type often have similar characteristics wrt. the two criteria:

