DSSim Results for OAEI 2008

Miklos Nagy, Maria Vargas-Vera, Piotr Stolarski and Enrico Motta
• Introduction and context
• DSSim Ontology mapping
• OAEI Mapping process
• Comments on OAEI tracks
• Future work
Introduction and context

- Question answering over multiple heterogeneous sources
- Incomplete and uncertain results of the different similarity algorithms
- Domain knowledge as a result of context dependent interpretation
- Distributed and dynamic environment
List all papers with keywords uncertain ontology mapping?

1. Answer composition component
2. Ontology mapping component
3. Broke agent
4. Mapping agent
5. Mapping agent
6. Mapping agent

WordNet

BibTeX/UMBC Ontology
BibTeX/MIT Ontology

Entry
Article
Resource
Publication
hasKeyword
• Concept, property names & hierarchies
• WordNet in order to exploit synonymy at the lexical-level
• Different syntactic similarity measures like Monger-Elkan, Jaccard
• Graph matching for semantic similarity
• Possible mappings as hypotheses in Dempster Shafer theory
Compound nouns

• Tokenise using different separators
• Consult background knowledge for each token
• Determine similarity for the intersection of tokens
• Assign compound noun into semantic rules
OAEI 2008 mapping process

Processor core 1

1. Divide ontologies into m*n fragments
2. Queue alignment jobs
3. Assign job to idle processor core

4. Ontology fragment
   - Create query graph fragment
   - Create ontology graph fragment
   - Assess syntactic and semantic similarity
   - Create and combine beliefs over similarities

Alignment Format

Producer alignments

Next iteration

Processor core 2

5. Ontology fragment
   - Create query graph fragment
   - Create ontology graph fragment
   - Assess syntactic and semantic similarity
   - Create and combine beliefs over similarities

Next iteration

Next iteration
Comments on OAEI tracks

Benchmark track

Pros

• Relatively small ontology sizes
• Most ontologies contain instances

Cons

• Relatively small number of real word ontologies
Anatomy track

Pros

• Relatively small ontology sizes
• Real world ontologies

Cons

• We believe that domain specific background knowledge is needed for good results i.e. biological, has to be confirmed ...
• Difficult names and labels for non experts (outer renal medulla peritubular capillary)
• No instances
Comments on OAEI tracks

Fao track

**Pros**
- Relatively small number of classes
- Large number of individuals
- Lot of additional information at the individual level

**Cons**
- Class information at the individual level
Comments on OAEI tracks

Directory track

Pros
• Small ontology sizes
• Large number of different ontologies

Cons
• Complexity for determining synonyms
  e.g. “News and Media”
• Individuals are modeled as classes
  e.g. 3D_Studio_Max
ML Directory

Pros

• Ontologies from different domains
• Manageable ontology sizes
• Large number of individuals compared to classes

Cons

• English-Japanese, Japanese-Japanese tests difficult to verify before submission
Library track

Pros

• Large real world ontologies
• Refer to broader and narrower concepts

Cons

• Difficult naming e.g. “gtt_291556558”
• SKOS -> OWL conversion, Chunk parsing
• Background knowledge in Dutch is necessary
Comments on OAEI tracks

Very Large cross lingual resources track

Pros

• Dutch-English/Dutch-English tests

Cons

• Difficult conversions from original format (DBPedia)
• Nearly unmanageable size
• Wordnet as background knowledge and source ontology
Conference track

Pros
• Large number of possible tests (around 200)
• Relatively small number of concept/properties
• Ontologies with different expressiveness

Cons
• Around 50% of possible alignments are empty
Future work

• Further test need to be carried out on fuzzy trust voting model
• Further optimise belief combination process
• Achieve better utilisation (query) of background knowledge
• OAEI 2009 participation...
Thank You!