Debugging is-a structure in ontologies

Patrick Lambrix, Valentina Ivanova, Zlatan Dragisic, Fang Wei-Kleiner
Defects in ontologies

- **Syntactic defects**
  - eg. wrong tags or incorrect format

- **Semantic defects**
  - eg. unsatisfiable concepts, incoherent and inconsistent ontologies

- **Modeling defects**
  - eg. wrong or missing relations
Example - incoherent ontology

- Example: DICE ontology
  - \( \text{Brain} \sqsubseteq \text{CentralNervousSystem} \sqcap \text{BodyPart} \sqcap \exists \text{systempart.NervousSystem} \sqcap \exists \text{region.HeadAndNeck} \sqcap \forall \text{region.HeadAndNeck} \)

  A brain is a central nervous system and a body part which has a system part that is a nervous system and that is in the head and neck region.

  - \( \text{CentralNervousSystem} \sqsubseteq \text{NervousSystem} \)

    A central nervous system is a nervous system.

  - \( \text{BodyPart} \sqsubseteq \neg \text{NervousSystem} \)

    Nothing can be at the same time a body part and a nervous system.

Slide from G. Qi
Example - missing is-a relations

- In 2008 Ontology Alignment Evaluation Initiative (OAEI) Anatomy track, task 4
  - Ontology MA : Adult Mouse Anatomy Dictionary (2744 concepts)
  - Ontology NCI-A : NCI Thesaurus - anatomy (3304 concepts)
  - 988 mappings between MA and NCI-A
    - 121 missing is-a relations in MA
    - 83 missing is-a relations in NCI-A
Influence of missing structure

- Ontology-based querying.

Medical Subject Headings (MeSH)

All MeSH Categories
- Diseases Category
  - Eye Diseases
    - Scleral Diseases
    - Scleritis

...
Influence of missing structure

- Incomplete results from ontology-based queries

 ![PubMed search result](image.png)

**Medical Subject Headings (MeSH)**

- All MeSH Categories
  - Diseases Category
    - Eye Diseases
      - Scleral Diseases
        - Scleritis

Return:
- 1363 articles
- 613 articles
- 55% results are missed!
Defects in ontologies

- Ontologies with defects, although often useful, also lead to problems when used in semantically-enabled applications.
  ➔ Wrong conclusions may be derived or valid conclusions may be missed.
Debugging the missing and wrong is-a structure of taxonomies
Outline

- Definitions
- Approach
- Experiments
- Conclusion
Outline

- Definitions
- Approach
- Experiments
- Conclusion
A taxonomy network consists of a set of taxonomies and sets of mappings between these taxonomies.
Defects in ontologies

- **Syntactic defects**
  - eg. wrong tags or incorrect format

- **Semantic defects**
  - eg. unsatisfiable concepts or inconsistent ontologies

- **Modeling defects**
  - eg. wrong or missing relations

→ **Solution requires domain knowledge.**
Assumptions and scope

- We focus on **taxonomies**, 
  - named concepts and *is-a* relations.
- We assume that all the existing mappings in the taxonomy network are **correct**.
- The mappings represent equivalence and subsumption.
Debugging is-a structure in taxonomy networks

Given a set of taxonomies networked by sets of correct mappings, how to detect and repair the missing and wrong is-a relations in these networked taxonomies?
Detecting missing is-a relations

- Domain expert – manual inspection
- Using external knowledge
  - Ontology learning
  - Discovery of subsumption relations (Hearst patterns, logical patterns)
- Using knowledge intrinsic to the network
Candidate missing is-a relations

Given two concepts A and B in a taxonomy O in the network. If “A is-a B” is logically derivable from the taxonomy network, but not from the taxonomy O alone, then “A is-a B” is a candidate missing is-a relation.

The candidate missing is-a relations need to be validated by a domain expert → wrong and missing is-a relations
Candidate missing is-a relations

- Two small pieces of MA and NCI-A, both about concept “joint”, and 3 equivalence mappings.
Repairing is-a relations

Repair the original taxonomies by
- adding a set of is-a relations to each taxonomy, such that the missing is-a relations can be derived from the extended taxonomy;
- removing a set of is-a relations from the taxonomies, such that the wrong is-a relations cannot be derived from the network

- Structural repair:
  - The is-a relations within the structural repair are called ‘repairing actions’.
Repairing missing is-a relations

Question:
How can we recognize structural repairs that are interesting for a domain expert?

→ **heuristics.**
Axiom-based Heuristic

Prefer to use structural repair **without non-contributing** repairing actions.
Information-based heuristic

Prefer to use structural repair with more informative repairing actions.

(limb_joint, joint) is more informative than (hip_joint, joint) and (elbow_joint, joint)
Strict hierarchy heuristic

Prefer to use structural repair which **does not change the existing is-a relations in the original ontology into equivalence relations.**

**(body part, joint)** will introduce an equivalence relation between ’joint’ and ’body part’.
Single relations heuristic

Assume that it is more likely that domain experts have missed a single relation than a chain of relations.

- Assume it is more likely that
  \((\text{ankle\_joint, limb\_joint})\)
  is missing than
  \((\text{ankle\_joint, } x_1) \text{ and } (x_1,x_2), \text{ and } ... \text{ and } (x_{k-1}, x_k) \text{ and } (x_k, \text{ limb\_joint}).\)
Repairing wrong is-a relations

- Find explanations (justifications)
- Remove part of the explanation
Outline

- Definitions
- Approach
- Experiments
- Conclusion
Overview of debugging approach
Phase 1: Detecting candidate missing is-a relations
Phase 2: Validating candidate missing is-a relations
Phase 3.1: Generating repairing actions for missing is-a relations
Example
Phase 3.2: Ranking missing is-a relations
Phase 3.3: Recommending repairing actions for missing is-a relations
Phase 3.4: Executing repairing actions for missing is-a relations
Repairing wrong is-a relations

- Phase 3.1: generate repairing actions
  - Based on justifications
- Phase 3.2: rank wrong is-a relations
  - Based on number of possible repairing actions
- Phase 3.3: recommend repairing actions
  - Based on occurrences in different derivation paths
- Phase 3.4: execute repairing actions
  - Compute consequences
Outline

- Definitions
- Approach
- Experiments
- Conclusion
Experiment “missing” - bib

- Bibliography dataset (2010 OAEI Benchmark)
Experiment “missing” - bib

Bibliography Dataset – 1 network

- Missing is-a relations
  - 22 in 101 (of which 12 redundant)
  - 1 in 301
  - 1 in 302
  - 1 in 303
  - 23 in 304 (of which 14 redundant)

- The whole debugging process took about 5 minutes.
Experiment “missing” - bib

- Bibliography Dataset – 4 small networks
  - Missing is-a relations
    - For 101-301: 1 for each ontology
    - For 101-302: 17 (of which 11 redundant) for 101 and 1 for 302
    - For 101-303: 1 for 303
    - For 101-304: 4 for 101 and 5 (of which 1 redundant) for 304
  - The whole debugging process took less than 5 minutes.

- Comparison 1 network / 4 networks
  - 301, 302, 303: same results in both scenarios
  - More missing is-a relations found and repaired in the scenario with 1 network
Experiment “missing” - Anatomy

Experiment on Anatomy dataset (2008 OAEI Anatomy)
MA: 2744 concepts, 1807 asserted is-a relations
NCI-A: 3304 concepts, 3761 asserted is-a relations
PA: 988 equivalence relations, 1 subsumption
→
new is-a relations: 205 for MA, 177 for NCI-A
total: 3 hours debugging time (almost all time on validation)
In most cases, the ranking and recommendations seemed useful.
Experiment “wrong and missing” - Anatomy

Experiment on Anatomy dataset (2010 OAEI Anatomy)
MA: 2744 concepts, 1807 asserted is-a relations
NCI-A: 3304 concepts, 3761 asserted is-a relations
PA: 986 equivalence relations, 1 subsumption

→

new is-a relations: 107 for MA, 64 for NCI-A
removed is-a relations: 3 from MA, 12 from NCI-A
total: 5 hours debugging time (almost all time on validation)
Outline

- Definitions
- Approach
- Experiments
- Conclusion
Extensions

- Taxonomies
  - Debugging wrong and missing is-a structure and mappings within networked taxonomies (WoDOOM12, ESWC13)
    - Experiment on Anatomy dataset (2010 OAEI Anatomy)
    - ToxOntology – MeSH (Swedish National Food Agency)
  - Aligning ontologies = detecting missing mappings (ESWC13)

- ALC acyclic terminologies (JIST12)

- Repairing missing is-a relations is an abduction problem (JIST12)
Future work

- Algorithms for more ontologies in more expressive languages
- Complexity of the abduction problem for different languages
- Preference criteria for solutions
References

- Lambrix P, Ivanova V, A unified approach for debugging is-a structure and mappings in networked taxonomies.
References


- WoDOOM – International Workshop on Debugging Ontologies and Ontology Mappings