

# Quality ontology engineering based on thesauri

Author **Daniel Kless**  
Supervisors Simon Milton, Edmund Kazmierczak, Jutta Lindenthal  
Studies/Stage PhD, close to submission  
Affiliation University of Melbourne  
E-Mail d.kless@student.unimelb.edu.au

## Aims and Objectives of the Research

The primary goal of my research is to understand the differences and commonalities of thesauri and ontologies. The insights gained are applied in a method for reengineering a thesaurus into an ontology.

## Justification for the Research Topic

Vocabularies (sometimes called terminologies) and ontologies do not appear to be very different from the perspective of a practitioner and are often used as if they are ontologies. Nevertheless, experts acknowledge that ontologies are different from vocabularies, but find it difficult to pin down the differences. The identification of the differences between ontologies and vocabularies is the main goal of my research. The differences are made explicit through a method that guides the re-engineering of a vocabulary into a *qualitatively good* ontology. My particular interest is in reengineering a thesaurus as a specific type of vocabulary.

While the reengineering method can facilitate the construction of complex ontologies by adapting existing thesauri, the insights into the differences between ontologies and vocabularies contribute to a better understanding, discrimination, construction and evaluation of ontologies as computer artefacts. They also reveal widespread misunderstandings of ontologies and the wrong application of logic-based languages like OWL, which has to be avoided, if ontology goals like the integration of knowledge shall be achieved in future.

## Research Questions

1. What are the differences and commonalities between thesauri and ontologies?
2. Which steps are necessary to reengineer a domain-specific thesaurus into an ontology?

## Research Methodology

Two methodically different steps can be distinguished in my research. First, the relations and relata in ontologies and the relationships and their relata in thesauri were systematically compared. In particular, thesaurus relationships and their relata as they are defined in the latest version of the international thesaurus standard ISO 25964 were analyzed against formally well-defined ontological relationships and relata from ontology literature—more specifically ontology literature based in realism.

The second step was a case study of reengineering a thesaurus into an ontology. More specifically, an excerpt of the AGROVOC thesaurus concerned with agricultural fertilizers was converted into an ontology. For this purpose we used the insights from the comparison of thesaurus and ontology as an input, but also applied top-level ontologies and adapted some best practices in ontology modelling advocated in the biomedical domain. From the case study we induced a general method for reengineering thesauri into ontologies.

## Research Results to Date

Our systematic comparison [2] made it clear that thesauri require structural and definitional re-engineering in order to be reused or treated as ontologies, but that adherence to the international standard for thesauri provides a good base for such reengineering. The relata as well as the relationships in thesauri need to be classified further before any matching to formal ontological relationships is possible. Isolated hierarchical relationships in thesauri then may correspond to the is-a relationship, specific mereological relationships, or fundamental relationships such as the instantiation between universals and individuals in ontologies. If such correspondences apply in domain-specific cases depends on whether the thesaurus relationships contribute to the specifications of *necessary and sufficient conditions* for their respective relata in the ontology—a function that relationships do not have in thesauri.

The comparison revealed that current “ontology” definitions lack a clarification that they must model necessary and ideally sufficient membership conditions of concepts only, but not, e.g., accidental, i.e., possible properties. This has led to the widespread misperception that data models, thesauri or other types of vocabularies are kinds of ontologies. Such thinking undermines the possibilities to integrate ontologies with each other and to reason over ontologies. Further, numerous ontology publications actually do not deal with ontologies and are wrongly classified to be about “ontologies”. In consequence there are also questionable statements about ontologies, e.g. with respect to their usefulness.

Our case study allowed us to distinguish eight steps that are necessary to reengineer a thesaurus into an ontology [1]:

1. Preparatory refinement and checking of the thesaurus
2. Syntactic conversion
3. Identification of membership conditions (in natural language)
4. Choice and alignment to top-level ontologies and formal relations
5. Formal specification of classes
6. Dissolving poly-hierarchies in the asserted ontology
7. De-coupling of independent entities
8. Adjustment of spelling, punctuation and other aspects of class and property labels

The sum of these steps shows that re-engineering thesauri ontologically requires far more than just a syntactic conversion into a formal language or other easily automatable steps. Steps 2, 6 and 8 may be at least partially automatable while the other steps appear to have no automation potential at the current state of the art. Steps 3-5 represent the core of ontological modelling in general and require considerable implementation effort if done qualitatively well [1]. Ontology quality should be measured by how *correct*, how *precise* and how *complete* the class membership conditions are specified, since it affects the correctness of the is-a hierarchy. Further investigation is required to determine when the effort of building qualitatively good ontologies is justified over using alternatives.

## References

1. Kless, D., Jansen, L., et al., (accepted and to appear 2012). A method for re-engineering a thesaurus into an ontology. In *Proceedings of the 7th International Conference (FOIS 2012)*. Formal Ontology in Information Systems. Graz, Austria: IOS Press.
2. Kless, D., Milton, S. & Kazmierczak, E. (under revision). Relationships and Relata in Ontologies and Thesauri: Differences and Similarities. *Applied Ontology*.
3. Kless, D., Lindenthal, J., et al., 2011. The difference between creating ontologies and applying SKOS and OWL. In A. Slavic & E. Civallo, eds. *Proceedings of the International UDC Seminar*. Formal approaches and access to knowledge: Classification & Ontology. The Hague, The Netherlands: Ergon Verlag, pp. 55–74.
4. Kless, D. & Milton, S., 2010. Towards Quality Measures for Evaluating Thesauri. In S. Sánchez-Alonso & I. N. Athanasiadis, eds. *Metadata and Semantic Research*. Communications in Computer and Information Science. 4th International Conference, MTSR 2010. Alcalá de Henares, Spain: Springer Berlin Heidelberg, pp. 312-319.
5. Kless, D. & Milton, S., 2010. Comparison of thesauri and ontologies from a semiotic perspective. In *Proceedings of the Sixth Australasian Ontology Workshop*. Conferences in Research and Practice in Information Technology. Advances in Ontologies. Adelaide, Australia: Australian Computer Society, pp. 35-44.