# A content-focused method for re-engineering thesauri into semantically adequate ontologies using OWL

Daniel Kless [a], Ludger Jansen [b], Simon Milton [a]
[a] Department of Computing and Information Systems, The University of Melbourne, Parkville, 3010 VIC, Australia
[b] Institute for Philosophy, The University of Rostock, August-Bebel-Straße 28, 18051 Rostock, Germany
Corresponding author: D. Kless, E-Mail: klessd@unimelb.edu.au (please CC: dkless@gmx.de)

**Abstract.** The re-engineering of vocabularies into ontologies can save considerable time in the development of ontologies. Current methods that guide the re-engineering of thesauri into ontologies often convert vocabularies merely syntactically and ignore problems arising from interpreting vocabularies as ontologies, i.e. as sets of statements of facts. Current re-engineering methods also do not make use of the semantic capabilities of formal languages in order to detect logical mistakes and improve vocabularies. In this paper, we introduce a content-focused method for building domain-specific ontologies based on a thesaurus, a popular type of vocabulary. Application of the method results in an ontology that not only adheres to the semantics of the description logic OWL, but also contains a semantically rich description of the modeled entities, enables non-trivial, automated reasoning, and can be integrated with other ontologies following the same development principles. We explain the motivation and sub-activities for each of the steps in our method and illustrate their application through a case study in the domain of agricultural fertilizers based on the ACROVOC Thesaurus. Our method shows, first and foremost, that a considerable manual effort is required to derive a semantically rich ontology from a thesaurus, particularly in connection with the alignment to a top-level ontology as well as for the identification and formal specification of membership conditions. Applying our method will likely change the structure of a thesaurus considerably. Our method is particularly useful where a highly reliable is-a hierarchy or consistent definitions are crucial.

Keywords: Thesaurus Re-engineering, Ontology development, OWL

## 1. Introduction

The creation of knowledge-dense ontologies can take tremendous time [1]. For this reason, it is desirable to re-use existing models as ontologies [2]. Also, the re-engineering of non-ontological models for their use as ontologies has become popular. Controlled vocabularies (referred to as "vocabularies" in the following), more recently known as knowledge organization systems and often referred to as terminologies, are examples of non-ontological resources and are generally considered interesting candidates for re-use as ontologies [3,4]. The reason is that such vo-cabularies have often matured over decades and contain several thousand or even hundreds of thousands of concepts and natural language terms. This elimi-nates or at least reduces the effort of eliciting con-cepts in the ontology development process. Second, the concepts in a vocabulary are generally structured through a number of relationships. These relation-ships can be used as a starting point for developing the structure of an ontology.

Today's practices of re-engineering vocabularies into ontologies differ significantly. The main reason is that there are different understandings of ontolo-gies. In this paper, we refer to ontologies as state-

ments of necessary and general features of a certain domain of reality in a computable formal language like *First Order Logic* (FOL) or one of the varieties of *Description Logics* (DL) [5]. Our specific focus in this paper is the re-engineering of vocabularies into ontologies described in the Web Ontology Language (OWL) [6] and adhering to the corresponding Description Logic semantics (OWL-DL) [7].

Although there are numerous ontologies that are coded in OWL-DL, for example the ontologies published by the OBO Foundry [8], we are not aware of any publication that describes an *instructive* and *holistic* method for re-engineering a vocabulary into such ontologies. Hahn [9] and Hahn and Schulz [10] give recommendations based on their experience with the ontological re-engineering of the UMLS meta-thesaurus; Wroe et al. [11] with regards to their re-engineering of the Gene Ontology. Nevertheless, these publications do not describe a method for the re-engineering of vocabularies into OWL-DL ontologies.

Cardillo et al. [12] developed a script that converts a thesaurus into an OWL representation without any further consideration of the correctness of the results or using the expressive potential of OWL to define the meaning of classes. The report of the NeOn project [13] does mention the re-engineering of a vocabulary into an OWL-DL ontology as "TBox re-engineering", but only refers to the Scarlet program [14] and the use of WordNet without detailing any exact procedure.

There are numerous general ontology engineering methods, some of which have the specific focus of developing an OWL-DL ontology, e.g., Borgida and Brachman [15] or Noy and McGuinness [16]. Moreover, there are content-focused methods that provide guidance for *specific aspects* of ontology engineering that are also applicable to developing an OWL-DL ontology, e.g., OntoClean [17], to improve the is-a hierarchy.

Our method aims not only at compliance to OWL standards, but also at developing *semantically adequate* ontologies that
   a) make full use of the semantic expressivity of OWL,
   b) are consistent and provide true reasoning results and
   c) can be integrated with other ontologies following the same development principles.
The method we will present guides specifically the re-engineering of a *thesaurus,* a specific type of vocabulary. Vocabularies come in a great variety, from simple lists of terms to thesauri, taxonomies or classi-

fication schemes [18]. These types can differ in their syntactic and semantic properties. A thesaurus is a type of controlled and structured vocabulary whose structural properties are well defined by international standardization efforts [19,20]. As there exist presumably several hundreds of thesauri that could be adopted as ontologies [21], our method can find a wide application. We will demonstrate the validity of our method by applying it to a portion of a specific thesaurus, namely the fertilizer branch of the AGROVOC Thesaurus [22].

The paper is structured as follows: In the subsequent section 2 we explain the distinctive characteristics of OWL DL-compliant ontologies and how they shape our re-engineering approach. Section 3 details how the re-engineering method was derived. Section 4 discusses the individual steps of our re-engineering method. In an earlier paper [23], we already provided an outline of the method. The present paper presents a matured version of the method in much more detail. In section 5 we will reflect on the method as a whole and summarize our results in section 6.

## 2. Characterization of the re-engineering method

Our goal is the re-engineering of a thesaurus into a semantically adequate OWL-DL ontology. In this section, we describe the essential properties of OWL-DL ontologies and contrast these with other approaches to pinpoint the distinguishing features of our re-engineering approach.

### 2.1. Characteristics of OWL-compliant engineering

The OWL syntax and the associated description logic semantics constitute a formal system that supports automated reasoning based on membership conditions and other features. Automated reasoning can be used both for automated consistency checks of ontologies (i.e. for proving the absence of contradictions) and inferring facts that have not explicitly been asserted [24]. Advantages of OWL-DL are its computational tractability, the reasoning support for consistency checking and for the generation of the inferred class hierarchy, as well as the use of XML-based syntaxes and unique identifiers (IRIs and URIs). These are also the reasons for our focus on OWL-DL ontologies in this paper.

A complete overview of the characteristics is given in the OWL specification [6]. Of particular importance are the strict separation of TBox and ABox, i.e., of classes and individuals, and the strict semantics of relationships.

OWL-DL standardly distinguishes between instances (individuals) and abstractions of them (classes). This separation corresponds to the distinction of the so-called TBox and ABox: The TBox, the 'terminology component', "contains intensional knowledge in the form of a terminology and is built through declarations that describe general properties of concepts". By contrast, the ABox, the 'assertion component', "contains extensional knowledge—also called assertional knowledge—knowledge that is specific to the individuals of the domain of discourse" [25]. In other words, the TBox concentrates on the intensional specification of classes using previously specified relationships while the ABox uses the definitions made in the TBox to describe particular things (individuals). The TBox acts, thus, as a metamodel for the ABox, "a model that consists of statements about models" [26]. We follow Guarino et al. [27] and Schulz and Jansen [28] in considering only intensional knowledge (the TBox) to be part of an ontology proper.

Concepts in thesauri are—with some exceptions—intensional entities that are labeled by general terms, terms that are predicableof more than one individual [29]. As is shown in figure 1, re-engineering thesauri into ontologies means that the majority of the thesaurus content (see (b) in figure 1) ends up in the TBox (2). Only few thesaurus concepts, in specific references to particular things such as the Mekong River or Rocky Mountains, end up in the ABox, but are then not considered part of the ontology (TBox).
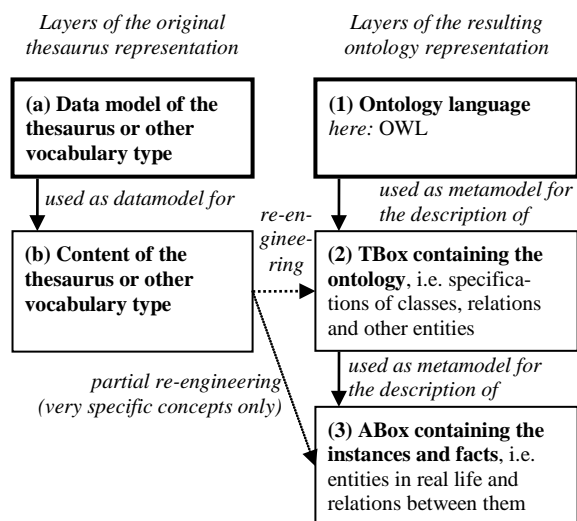


Figure 1. TBox re-engineering process for thesauri and other types of vocabularies

A thesaurus uses a data model (see (a) in figure 1) to determine its logical structure and organize its content [30]. Shifting the content of the thesaurus into the TBox requires structural re-engineering that is caused by the differences between the thesaurus data model (a) and the metamodel that underlies the formal system and, thus, the ontology language (see (1) in Figure 1).

A specific use of relationships is another significant characteristic of OWL-compliant ontologies. When using relationships in the TBox of such ontologies, one needs to respect the following three characteristics:

(1) Relationships involve implicit or (in OWL) explicit quantification, which is relevant for the semantics of relational expressions [31]. Statements like *'bow' has part 'string'* or *'quality' has bearer 'material object'* are, thus, not well-formed in OWL, but require the insertion of quantifiers like *'some'* or *'only'*

(2) Any relationship from a class *A* to another class *B* needs to apply to each instance of *A* and has, thus, the logical force of a necessary membership condition for the class *A*. For this reason, a statement like *'overgrazing' causes some 'desertification'* would not be proper ontological content because the process of overgrazing does not *necessarily* cause desertification: There are instances of overgrazing that do not lead to desertification.

(3) This implies that the relationship *'A isRelatedTo some B'* does not normally imply the inverse relationship *'B hasRelationFrom some A'*. E.g., every bow has as part some bow string, but not every bow string is part of some bow. Even if a relationship is (as allowed by OWL) explicitly defined to be the inverse of another relationship, this does not necessarily carry over to the class level. For example, *'human' has part some 'head'* is a true statement, but not so *'head' part of some 'human'*.

In OWL, these rules also apply to the ABox, with the difference being that the relationships describe characteristics of the instances of classes, not membership conditions. The rules apply for all relationships except for the subclass relationship and the instantiation relationship, which are built-in features of OWL-DL.

OWL brings with it certain limitations, which are often discussed as *expressivity limitations* [32,33]. Here we want to list some rather macroscopic problems of OWL; some of them are typical for any at-

tempt to use formal logics to describe the meaning of natural language statements [34]:

- OWL is limited to countable quantifiers (*all*, *some*, *min* x, *max* y). There are no vague or proportional quantifiers (e.g., many, sufficient, most, nearly); hence statements like "snow is mostly white" cannot be expressed in OWL.

- Unlike some forms of modal logic, OWL has no primitives that could express the modality of a statement, i.e. a statement cannot be qualified by through modal operators such as "It is *usually/typically/possibly/necessarily* the case that", "It is *likely/forbidden/desired* that" or intensional contexts like "X *thinks/believes/is certain/supposes* that".

- OWL has no primitives that can express the tense or aspect of a statement, e.g., statements like "John *was/is/will be* rich" are not possible. It cannot be indicated *when* or *under what circumstances* a certain statement was given or when it will be true. Thus, only timeless statements can be expressed in OWL.

- As far as the definition of general terms through classes is concerned, OWL can only provide statements that are true for all members of the class, not just some members; i.e. the (true) statement "Some fertilizers pollute soil" cannot be expressed, only the (false) "All fertilizers pollute soil".

- As OWL has no variables, it has peculiar problems with nested quantification and identity statements.

These limitations are particularly significant when comparing ontologies described in OWL with thesauri and hence represent problems for any project of re-engineering a thesaurus into an ontology, including the present case study.

## 2.2. Comparison with other re-engineering approaches

In Knowledge Engineering, the term "ontology" is often used in a loose sense to describe different types of models that do not share the characteristics of OWL-DL ontologies which we described in the preceding section. There are at least two groups of approaches in the re-engineering literature according to the respective 'loose' sense these papers connect with the term "ontology".

A first group of re-engineering publications does not refer to any specific description language of ontologies, but rather associates ontologies with the freedom of defining customized relationships between concepts, which is not normally the case with thesauri and other types of vocabularies. This re-engineering approach, thus, focuses on changing the content of a given thesaurus, mainly through a refinement of relationships without changing the overall structure of a thesaurus. An example is Soergel et al. [35], whose approach also underlies the publications of Kawtrakul et al. [36] and Sánchez-Alonso and Sicilia [37] and has similarities with ontological augmenting of thesaurus relationships described by Tudhope et al. [38].

In a second group of re-engineering publications, an ontology is a description of a model that uses the Resource Description Framework (RDF) [39] and adheres to the RDFS Semantics [40]. According to this approach, as displayed in figure 2, re-engineering a thesaurus or other vocabulary type into an ontology means, first, to describe the data model of a thesaurus or other vocabulary types (a) in a schema (2) using RDFS (1). In the case of thesauri it can suffice to adopt the Simple Knowledge Organization System (SKOS) [41], which is closely oriented on the thesaurus data model described in ISO 25964-1:2011 [20]. The schema is then used to describe the content of a domain-specific thesaurus or vocabulary (b), which does not require any structural changes in the thesaurus or vocabulary [13] and, thus, does not check the adequacy of the content. This kind of re-engineering is, therefore, easy to automate and re quires relatively
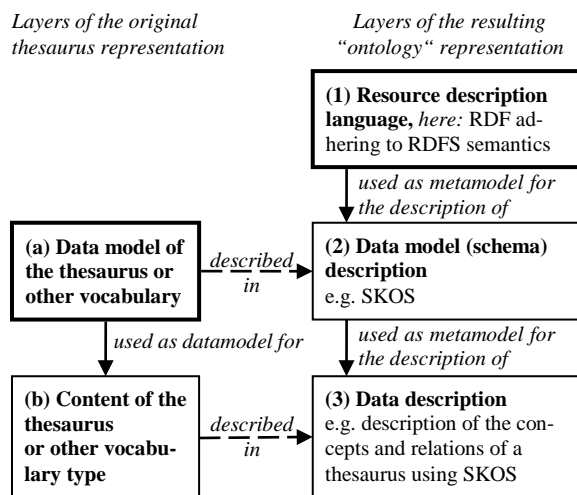


Figure 2. ABox "re-engineering" process for thesauri and other types of vocabularies

little manual work. Such re-engineering is described by Villazón-Terrazas [4] and used by the NeOn project [13] and van Assem [3]. Both Villazón-Terrazas and the NeOn project use the terms "TBox re-engineering" and "ABox re-engineering" in ways that do not comply with DL semantics.

Neither of the two groups of re-engineering approaches separates classes from individuals. E.g., in RDFS, instances of classes can themselves be classes. A datatype (rdfs:Datatype), for example, is both an instance (rdf:type) of a class (rdfs:class) and a sub-class of such class (rdf:SubClassOf) . Moreover, instances can have instances, which is not possible in OWL-DL. For example, the sub-class relationship (rdf:SubClassOf) is declared an instance (rdf:type) of a property (rdf:Property) and a property (rdf:Property) is declared an instance (rdf:type) of a class (rdfs:class) [40].

Moreover, neither of the two approaches recognizes the importance of defining necessary membership conditions of classes. As a result, the ontologies produced by these re-engineering approaches cannot be used as OWL-DL ontologies (while it is not difficult to represent OWL-DL ontologies in RDFS using the mappings from the OWL specification [41]). Thus far, the re-engineering strategy presented here differs significantly from these two groups of approaches.

## 3. Elaboration of the re-engineering method

The re-engineering method that we present in this paper was developed in two phases: We started with (1) developing a naïve re-engineering method based on previous literature and then (2) refined and validated the method during the case study. In the first phase, we compared the structure of thesauri with the structure of ontologies theoretically. More specifically, we compared the thesaurus structure described in the thesaurus standard ISO 25964-1:2011 [20] with the structure of realist ontologies [42] and their specific representation in the description logic OWL [43,44]. Based on this structural comparison, we translated the identified differences and similarities into an initial set of steps for re-engineering thesauri into ontologies.

Additionally, we elicited certain steps for the general development or engineering of semantically adequate ontologies from the literature. However, we did not find any single method comprising all the steps that we have adopted. The combination of the steps from the theoretical analysis and the general ontology

engineering literature constituted the naïve re-engineering method and is laid out in Appendix 1.

In the second phase of refining and validation, we applied the naïve re-engineering method in a case study in order to re-engineer a portion of an existing thesaurus into a semantically adequate ontology. In this course, we added, merged or removed certain steps, changed their sequence and introduced sub-activities. Appendix 1 provides an overview of the changes by showing how the steps of the naïve re-engineering method are related to the steps in the final re-engineering method that we will introduce in the following section.

During re-engineering we were confronted with two challenges. First, re-engineering was highly time-consuming, which was anticipated. This challenge limited the number of representational units that could be feasibly re-engineered in the case study. In a real-world scenario, time is, of course, correlated with costs. Second, a variety of skills are required for the re-engineering that is rarely concentrated in a single person: knowledge of the structure of thesauri, experience in logic-based modeling (here: experience in the correct use of the modeling language OWL), familiarity with an appropriate modeling tool, knowledge of specific philosophical notions, familiarity with specific existing top-level and domain-specific ontologies, but also knowledge in the domain of the thesaurus to be re-engineered (here: agriculture). This challenge we met by working in a team to cover the required skills.

For the case study we chose the fertilizer branch of the AGROVOC thesaurus [22], which comprises 31 concepts subordinated to 'Fertilizers'. Additionally, we re-engineered a number of other concepts from the AGROVOC thesaurus that are closely related to fertilizers and were frequently needed when defining membership conditions of fertilizer types (step 3 of our method) and formalizing these (step 5), for example 'plant nutrient'. We chose the fertilizer-related portion of the AGROVOC thesaurus because of the specific interest in a project participant in a fertilizer ontology and because the AGROVOC is a mature and widely used thesaurus.

## 4. The re-engineering method and its application in a case study

Our re-engineering method consists of seven steps that are shown in figure 3. In the figure, the big arrows connecting the steps indicate that the method is expected to be applied sequentially, except for steps 3

5

```
┌─────────────────────────────────────────┐
│ 1. Preparatory refinement and checking   │
│    of the thesaurus                       │
└─────────────────────────────────────────┘
                    │
                    ▼
┌─────────────────────────────────────────┐
│ 2. Syntactic conversion                  │
└─────────────────────────────────────────┘
                    │
                    ▼
┌─────────────────────────────────────────┐
│ ┌───────────────────────────────────┐   │
│ │ 3. Alignment to a top-level       │   │
│ │    ontology                        │   │
│ └───────────────────────────────────┘   │
│ ┌───────────────────────────────────┐   │
│ │ 4. Identification of membership   │   │
│ │    conditions                      │   │
│ └───────────────────────────────────┘   │
│ ┌───────────────────────────────────┐   │
│ │ 5. Formal specification of        │   │
│ │    membership conditions           │   │
│ └───────────────────────────────────┘   │
│ ┌───────────────────────────────────┐   │
│ │ 6. Adjustment of spelling,        │   │
│ │    punctuation and other           │   │
│ │    aspects of entity labels        │   │
│ └───────────────────────────────────┘   │
└─────────────────────────────────────────┘
                    │
                    ▼
┌─────────────────────────────────────────┐
│ 7. Dissolving poly-hierarchies           │
└─────────────────────────────────────────┘
```
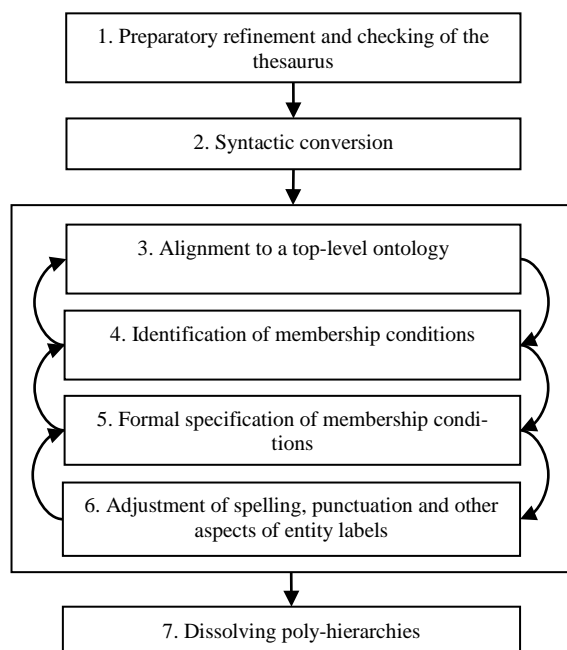
Figure 3. Method for engineering quality ontologies based on thesauri

to 6 that form a block of strongly interdependent and, thus, iteratively applied steps. Appendix 2 provides a more detailed overview of the method by summarizing the sub-activities for each step. The following subsections will, for each of the steps, discuss the purpose, provide an explanation of the activities involved, demonstrate the step to re-engineer the chosen portion of the AGROVOC thesaurus, and, finally, discuss the respective step. The demonstration of each step is structured according to the sub-activities that we will introduce in the explanation of the step.

By applying the re-engineering method, the thesaurus gradually becomes an ontology. We will refer to this model-in-transition as the "re-engineered thesaurus" or the "emerging ontology" to avoid confusion. Steps 3, 5, 6 and 7 will likely be realized using an ontology editor like Protégé [45].

## 4.1. Step 1: Preparatory refinement and checking of the thesaurus

### Purpose

We base our re-engineering method on the thesaurus standard ISO 25964-1:2011 [20]. However, thesauri are not necessarily in line with this particular standard: thesaurus standards have been developed and changed over time, whereas the data structure of an actual thesaurus system is practically inert after it

has been implemented. Hence it is possible or even likely that domain-specific thesauri may often not have adopted all the changes in the standards and re-engineering should begin with checking and refining the thesaurus so that further steps can rely on a stable basis. Furthermore, some of the optional features of thesauri-like node labels for indicating characteristics of division of the thesaurus concepts can be helpful for later analytical steps; for this reason, we encourage their use at this point.

In some cases, the refinement of the thesaurus may be impeded by the specific thesaurus management software used. For this reason, this step may be customized, combined with other steps or even skipped if the specific case of the re-engineered thesaurus requires or allows doing so. Nevertheless, various activities of this step are pivotal for deriving a useful basis for the is-a hierarchy of an ontology.

### Actions to be taken

In accordance with the ISO thesaurus standard ISO 25964-1:2011, a thesaurus should possess the following features, and we will now discuss the matching sub-activities to ensure these features:

a. Distinction between concepts and terms
b. Distinction between different types of hierarchical relationships
c. Rejection of invalid relationships
d. Removing hierarchical cycles
e. Assigning orphans to the thesaurus hierarchy
f. Identification of arrays of concepts based on common characteristics of division.

(a) The distinction between **concepts**, "units of thought" [20], and **terms**, "words or phrases used to label a concept" [20], is explicit in the data model in the thesaurus standard ISO 25964-1:2011. If a thesaurus does not make this distinction, then concepts need to be created that represent the preferred terms and their respective bundle of non-preferred terms. Eventual corrections should generally be automatable. Hierarchical or associative relationships relating one or two non-preferred terms are considered erroneous in term-based thesauri and should be represented as concept-to-concept relationships. Similarly, if there are relationships between preferred terms, these should be represented as relationships between the concepts that these terms label. Definitions and other notes that concern the concept as a whole should be transferred from the terms to the concept.

(b) **Hierarchical relationships** in thesauri summarize a variety of ontologically distinct relationships

6

that may or may not be distinguished explicitly: (1) the **generic relationship**, "the link between a class or category and its members or species" (e.g., 'birds' and 'parrots'), (2) the **hierarchical whole-part relationship**, which is correctly applied if the part belongs *uniquely* to the whole (e.g., 'bicycle wheel' and 'bicycle') and (3) the **instance relationships** between a general concept and an instance (e.g., 'Mountains' and 'Alps') [20]. For the purpose of re-engineering a thesaurus into an ontology, these kinds of hierarchical relationships *must* be distinguished explicitly.

(c) In the course of differentiating hierarchical relationships, relationships that fail to conform to the semantics of "relationship", as defined in the thesaurus standards, should not be represented in the emerging ontology. Also, **associative relationships,** used for "suggesting additional or alternative concepts for use in indexing or retrieval" [20], can be ignored at this stage; these are to be applied between "semantically or conceptually" related concepts that are not hierarchically related [20]. Their usefulness in ontologies will be critically assessed in step 5.

(d) The thesaurus should also be analyzed for **cyclic hierarchical relationships**. Such cycles are considered erroneous in thesauri and would introduce a logical contradiction in the emerging ontology. Cycles are best addressed in connection with step 3 of our method.

(e) **Orphans**, i.e. concepts that are not hierarchically connected to any other concepts, may occur if the thesaurus management software does not check for their occurrence when deleting or entering concepts during the maintenance of a thesaurus. They would appear as unwanted top-level classes in the ontology and, thus, need to be assigned an appropriate place in the hierarchy. Alternatively, the term representing the concept can be assigned as a non-preferred term to an existing concept in the thesaurus.

(f) For later steps in the re-engineering method it is worth introducing **node labels** to form **thesaurus arrays** where different **characteristics of division** can be identified. For example, the node label 'by location' indicates the location as a common characteristic of division for the concepts 'ground water' and 'surface water' and can be used to group them in a thesaurus array. While there is some guidance for this kind of "facet analysis" for the identification of node labels [46,47], it remains an intellectual activity, for which no proper mechanical procedure is available.

Thesauri may contain further kinds of errors such as unidirectional relationships between concepts, different thesaurus relationships between the same pair

of concepts, terms with exactly the same spelling assigned to different concepts, or hierarchical or associative relationships between non-preferred terms in term-based thesauri. Such errors may become the source of populating structural problems in thesauri that may be difficult to resolve later. They also result in mistakes when adopted in the ontology and should be detected by thesaurus management software [20]. We will not further discuss such errors here.

*Application of the step to the fertilizer ontology*

(a) The AGROVOC thesaurus does not distinguish between concepts and terms. Unique identifiers (term codes) are provided for terms, but not for concepts. Figure 4 shows how the thesaurus terms have been transformed to be compatible with the concept-based thesaurus structure recommended in ISO 25964-1:2011. While non-preferred terms point to a preferred term in the original term-based thesaurus, a concept is introduced for every preferred term when changing to a concept-based thesaurus. The preferred term and the non-preferred terms point to the concept in a concept-based thesaurus and their status as either preferred or non-preferred terms is indicated through different relationships or in meta-information about a term. The described separation between terms and concepts did not require a distinct effort, but could be realized implicitly in the course of the naïve conversion (step 2).

(b) Like many other thesauri, AGROVOC does not distinguish between different types of hierarchical relationships. However, our analysis revealed that all
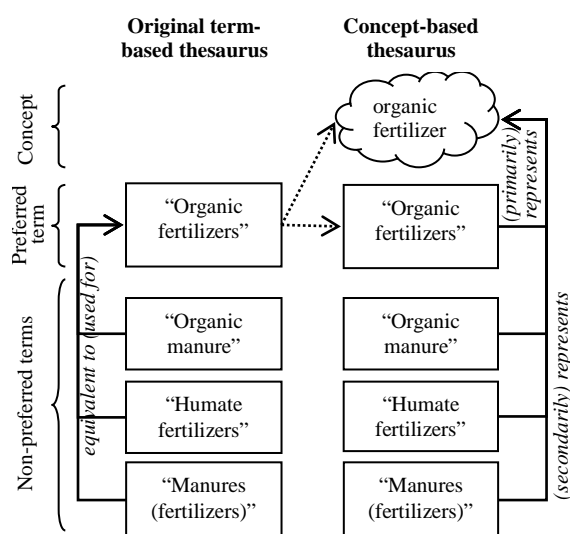


Figure 4. Conversion process from a term-based thesaurus like the AGROVOC to a concept-based thesaurus

hierarchical relationships between 'fertilizer' and its subordinated concepts are, by chance, proper generic relations between super-concepts and sub-concepts, as shown in figure 5. Other parts of the AGROVOC thesaurus do also display the other types of hierarchical thesaurus relationships like the instance relationship (Colorado River—Rivers) or the hierarchical part-of relationship (Root hairs—Roots).

```
Fertilizers
    Nitrogen fertilizers
        ammonium fertilizers
        nitrate fertilizers
    Phosphate fertilizers
        Rock phosphate
        Superphosphate
    Potash fertilizers
    Calcium fertilizers
    Magnesium fertilizers
    Sulphur fertilizers
    Compound fertilizers
        NPK fertilizers
        Nitrogen phosphorus fertilizers
            Nitrophosphates
        Nitrogen potassium fertilizers
        Phosphorus potassium fertilizers
    Micronutrient fertilizers
    Organic fertilizers
        Biofertilizer
        Composts
            Potting composts
        Fish manures
        Green manures
        Guano
    Organomineral fertilizers
    fertilizer combination
        fertilizer pesticide combinations
    Inorganic fertilizers
    Liquid fertilizers
    Liquid gas fertilizers
    Slow release fertilizers
```

Figure 5. Concept hierarchy in the re-engineered thesaurus. (Capitalization follows AGROVOC.)

(c) We noted some erroneous relationships amongst the fertilizer-related concepts. Specifically, some concepts were hierarchically related and associated at the same time. For example, 'Biofertilizers' was not only associated with 'Fertilizers', but also hierarchically subordinated to 'Fertilizers' (along the path of 'Organic fertilizers'). The erroneous associative relationships were simply ignored in our case study because, as we will argue in subsection 4.2, they will not be transferred into the ontology. We did not encounter relationships using a non-preferred term as a relatum that we would have to consider as a structural relationship in the ontology. We found only one scope note connected to a non-preferred term; it could easily be assigned to the concept itself because

there was no competing scope note for the preferred term.

(d–e) We could not detect any *hierarchical cycles* or *orphans* in the hierarchy.

(f) The AGROVOC thesaurus does not contain any *node labels* indicating characteristics of division. There are, however, several characteristics that can be used to group fertilizer such as the type of dominating plant nutrient, the number of plant nutrients, or the release time of plant nutrients. The complete list of the arrays defined by us with their respective node labels is provided in Appendix 3.

Our analysis revealed that checking and refinement of a thesaurus against standards is necessary to ensure a reliable basis for subsequent steps of the re-engineering process. At this stage, the fertilizer-related part of the AGROVOC thesaurus now conforms to the ISO standard.

*4.2. Step 2: Naïve conversion*

*Purpose*

Naïve conversion aims at a representation of the thesaurus in the formal language OWL-DL, a standard format that allows an unambiguous interpretation of the emerging ontology. Moreover, the formal representation in OWL-DL allows using automated reasoning tools to check the ontology for consistency (the absence of contradictions from the joint assertions made in an ontology), and infer the full class hierarchy in later steps.

Because of some fundamental structural differences between the thesaurus data model and the structural specification of OWL-DL, we connect the syntactic conversion with some initial structural changes of the thesaurus. Nevertheless, at this stage these, changes are implemented naïvely or mechanically only. It is, thus, possible that the model resulting from the syntactic conversion shows inconsistencies and contradictions that can later be detected through automated reasoning. The correction of these inconsistencies and contradictions is the subject of the following re-engineering steps.

*Actions to be taken*

In this step, two actions are to be applied sequentially:
    a.   Choice or development of conversion tools
    b.   Conversion of the thesaurus into the formal language

(a) It is desirable to carry out the described syntactic conversion automatically with conversion tools,

particularly when the goal is to re-engineer a complete thesaurus. The likelihood of being able to use existing tools, e.g., the one by Cardillo et al. [12], instead of needing to write customized scripts or programs is higher if the thesaurus is available in common exchange formats such as SKOS [48].

(b) After the refinement in step 1, the thesaurus is assumed to be concept-based according to ISO 25964-1:2011. On this basis, we can convert the thesaurus syntactically into a representation through a formal language by applying the mappings between representational units in thesauri and OWL as shown in figure 6. The diagram is to be read as follows: some concepts (in thesauri) reference individuals (in OWL). The name of the relation (*in italics*) expresses the meaning of the relation in the direction indicated.
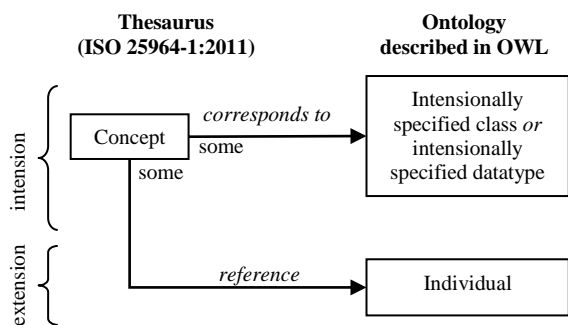


Figure 6. Relatedness of the relata in thesauri and the relata in OWL

Thesaurus concepts and facets in the role of top-level elements, can either correspond to an intensionally specified class or an intensionally specified datatype. In some cases, thesaurus concepts do not correspond to intensionally specified classes, but rather refer to individuals (e.g., the Yangtze River) or a particular collection of individuals (e.g., the Rocky Mountains as a particular collection of mountains). When defining the classes or the individuals in the emerging ontology, it is advisable to adopt RFC 3986 [49] or other conventions for the names of the entities (the identifiers, called URIs and IRIs in OWL).

The terms of a thesaurus and the labels of the facets now become labels of classes. Language tags can be used to distinguish the languages of the labels. Subtypes of labels need to be defined if it is desired to keep the distinction between preferred and non-preferred terms. Definitions, scope notes, and other notes and housekeeping information can be transferred to comments or appropriately defined subproperties thereof. It might also be desirable to transform node labels into **housekeeping classes** that serve on-

tology maintenance and navigation purposes, although they do not match any proper feature in the domain to be modeled. For example, we could, according to the material collected in Appendix 3, introduce classes labeled "Fertilizer by type of dominating plant nutrient" or "Fertilizer by amount needed by plants". It should be clear that these classes do not differ in their extension; they are, in fact, equivalent with the class 'fertilizer'. This equivalence, however, is weakened to a subclass-relationship in order to artificially make these nodes and the partitions represented by them distinguishable. Such housekeeping classes are needed as a workaround because OWL does not provide a modeling primitive corresponding to node labels.

Figure 7 shows mapping for relationships using the same notation. The **generic relationships**, which often dominate over the other kinds of hierarchical relationships in thesauri, are adopted as is-a relationships in ontologies, which are stated by a subclass axiom. In principle, a generic relationship could also be modeled as a data sub-property axiom in OWL, but we rather advise not to use data types in ontologies to avoid multiple hierarchies [28]. Note that at this stage the is-a relationships are preliminary only; they can become a subject of smaller or more fundamental changes in connection with steps 3 to 5.
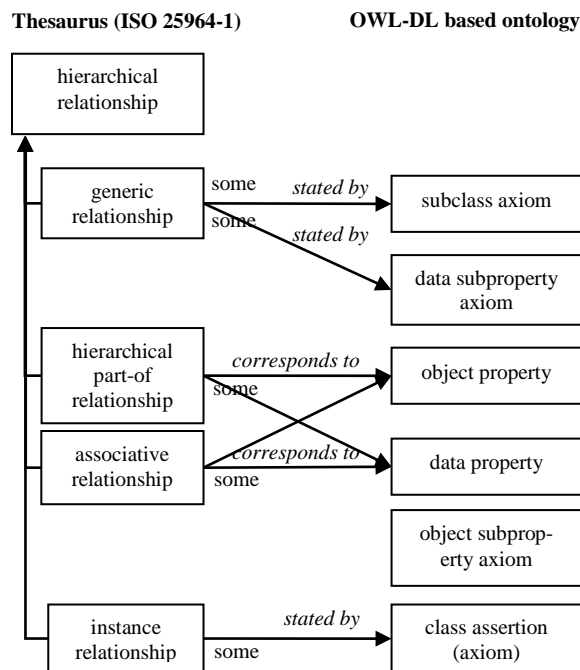


Figure 7. Relatedness of relationships in thesauri and relationships in OWL

**Hierarchical whole-part relationships** in thesauri should be tentatively modeled as unspecific has-part or, less commonly, as part-of relationships and represented by object properties or (less commonly) data properties in OWL. The relationships may be subject to further refinement depending on the set of formally defined relationships that shall be adopted (see step 5). Moreover, the hierarchical whole-part relationships as well as other relationships are subject to validity assessment in step 5, i.e. they must be membership conditions of the classes that they connect.

Correct **instance relationships** in thesauri correspond to relationships between an individual and a class—an assertion that is generally not considered part of the ontology. It rather belongs to a knowledge base, which can also be represented in OWL; in this case, instance relationships from a thesaurus are expressed by class assertion axioms in OWL.

**Associative relationships** may hint at relations that may contribute to the formal specification of ontology classes. We recommend checking the usefulness of associative relationships after step 5 rather than converting them directly into relationships in the ontology during the present step. In order to be relevant for the ontology, associative relationships need to be matched to appropriate ontological relationships, be represented by object properties in OWL (or, less commonly and not recommended by us, by data properties), and, just like the hierarchical whole-part relationships, must be membership conditions of one or both of the classes that they connect in order to be validly applied in the ontology. In our case study, no associative relation proved to be a valid ontological relationship.

OWL allows ordering relations hierarchically by means of subproperty axioms. Such an axiom expresses an is-a relationship between two relations, e.g., the structural-part-of relation can be said to be a subproperty of the part-of relation. Such hierarchies of relation are, however, not typical for thesauri.

*Application of the step to the fertilizer ontology*

An automatic syntactic conversion was not worth the effort in our re-engineering case, first, because of the particular export formats, second, because of the small portion of the thesaurus that we actually wanted to target. Although the AGROVOC website offers an OWL version of the AGROVOC thesaurus, this file has (1) computing problems as well as (2) structural problems:

(1) With a size of about 400 Megabytes, the file was far too large to be processed efficiently. It re-quired a computer with 8 processing cores and 8 GB of free memory to load the file in less than an hour. We know of no programs that support splitting ontology files of such a size into smaller portions.

(2) The way the OWL file is structured is not useful for our purpose. Most classes are direct siblings of the top concept "Thing" and very few classes are subordinated by the subclass axiom. We wanted to start with the class hierarchy mirroring that of the original AGROVOC thesaurus. An even bigger problem is that the class labels were not attached to the classes in a way that Protégé could display them.

For these reasons, and since we wanted to re-engineer a relatively small portion of the AGROVOC thesaurus only, it was faster for us to enter the class hierarchy for the 'fertilizer' tree manually using the Protégé-OWL editor. We started the conversion by creating classes for all fertilizer concepts. We decided not to introduce any arrays or household nodes into the ontology.

In a second step, we added the terms as labels to the classes. We retained the distinction between preferred and non-preferred terms by assigning them to the annotation properties "preferred term" and "non-preferred term" respectively and by introducing a subproperty of the default property "label". We also copied the preferred term to the "label" annotation property for further modification (see step 6). Furthermore, we defined a "scope note" as a subproperty of the default "comment" annotation property and copied the scope notes for the concepts into this field. Terms and notes in languages other than English were omitted when entering the thesaurus terms as class labels.

The fertilizer branch of the AGROVOC thesaurus contained generic relationships only. We adopted these as is-a relationships in our fertilizer ontology, though they are subject to further validation in later steps.

*4.3. Step 3: Alignment to a top-level ontology*

*Purpose*

Since the thesaurus hierarchy may have been a mix of generic relationships and other kinds of relationships, there may be several disconnected portions of classes organized through is-a relationships after the naïve conversion in step 2. Leaving things in this state would often mean that the upper-most classes of these portions are all assigned to the most general class, i.e. to the class 'Thing' (as it is called in Protégé). It is the goal of the present step to organize these

portions into a coherent is-a hierarchy and, thus, to connect them—directly or indirectly— by aligning them to a common top-level ontology.

Alignment to top-level classes is of considerable importance, for it is meant to establish interoperability with other ontologies aligned to the same top-level ontology. It also allows an economic specification of membership conditions of classes because subordinate classes inherit membership conditions and other formal specifications from superordinate classes (e.g., disjointedness from other classes) through is-a relationships. The alignment also establishes connections in the emerging ontology that are required for nontrivial automated reasoning.

*Actions to be taken*

The alignment comprises the following sequence of activities:

  a. Choice of an existing top-level ontology
  b. Alignment of the separated portions of the emerging ontology to the top-level ontology

(a) The choice of an existing top-level ontology is the most fundamental step. It involves getting an overview of existing top-level ontologies. Some of the commonly cited top-level ontologies include the *Descriptive Ontology for Linguistic and Cognitive Engineering* (DOLCE) [50], the *Basic Formal Ontology* (BFO) [51], the *General Formal Ontology* (GFO) [52] or the upper levels of *CyC* [53]. Further top-level ontologies are the *Suggested Upper Merged Ontology* (SUMO) [54] or *Yet Another More Advanced Top-level Ontology* (YAMATO) [55]. These top-level ontologies are generally published in OWL. Borgo and Vieu [56] give a brief introduction to most of them. Mascardi, Cordì and Rosso [57] performed a meta-analysis of comparisons of top-level ontologies. Semy, Pulvermacher and Obrst [58] list dimensions for comparing top-level ontologies, but specifically discuss SUMO, Upper Cyc, and DOLCE only. It should be noted that the upper-level hierarchy of Cyc is often not considered to be a proper foundational ontology at all [56], but rather a result of many historically explicable twists and turns [53,59] that has been comprehensively criticized [60].

(b) Aligning disconnected portions of the emerging ontology to the chosen top-level ontology aims for a state where every such portion of hierarchically connected classes is either directly or indirectly (through the class hierarchy) connected to the top-level ontology without creating any hierarchical cycles. To achieve this situation, each top-level class of a thesaurus portion is asserted to be *equivalent to* or *subclass of* either (1) a class of the adopted top-level ontology itself, (2) a class of another domain-specific ontology that is aligned to the top-level ontology (explained in step 4.a) or (3) a class of another portion of our emerging ontology that is aligned to the top-level ontology. The newly defined is-a relationships resulting from the alignment as well as the generic relationships that were adopted as *prima facie* candidates for is-a relationships in step 2 of our method (the is-a relationships *within* the portions of classes that we connected in this activity) will be checked and possibly changed in step 5 of our method. In the current step, it can already be checked whether membership conditions of the top-level classes apply, respectively, to all former thesaurus concepts subsumed under them.

Special consideration should be given to poly-hierarchies, i.e. to classes with more than one parent in the is-a hierarchy. As described in [61] and [62], poly-hierarchies do often come with conflicting membership conditions inherited from the various hierarchical paths. Thus, the existence of poly-hierarchies frequently indicates mistakes in the is-a hierarchy. Non-conflicting poly-hierarchies (also referred to as "multiple inheritance" or "diamond hierarchies") [63] are addressed in step 7.

*Application of the step to the fertilizer ontology*

(a) While most top-level ontologies are domain-independent, there are also so-called upper-domain ontologies that describe general kinds of certain domains. Since fertilizers belong to the field of biochemistry, we decided to use BioTop [64], an upper-domain ontology for the life sciences. BioTop is particularly suited for our purposes because it provides (1) a fine-grained distinction of material entities, (2) a comprehensive set of formally defined relationships [65], and (3) bridges to the most common top-level ontologies in the life sciences, i.e. BFO and DOLCE. As a result, our re-engineered fragment of AGROVOC can be used in combination with either of these two top-level ontologies.

(b) Our naïve conversion in step 2 of our method resulted in a single portion of hierarchically connected classes. We decided to align the class 'fertilizer' to be a subclass of the BioTop class 'compound of collective material entities'. A collective material entity is an aggregate of material entities that belong to the same type [63]. But as fertilizers are mixtures of different kinds of molecules, we decided to model them as compounds of collective material entities. In fact, the thesaurus was designed to describe real-life ferti-

lizers that can be purchased as products, and these will never be the pure substances.

*Discussion*

While some authors have doubts about the utility of top-level ontologies [66], our experience with the alignment is that top-level ontologies have an important guiding function by asking us to make categorial distinctions and decisions. BioTop, in particular, presented itself as a bundle of helpful micro-theories about ontological problems, for example, the distinction between dependent and independent entities. Thus, the categorial distinctions built into BioTop reduced the burden of decision-making and prevented misclassifications.

Naturally, adopting a top-level ontology implies a commitment to the specific theories that underlie the distinctions of the categories and relations. Even without weighing the advantages and disadvantages of BioTop against potential alternatives (e.g., adopting DOLCE [67] or BFO [51] and the Relation Ontology [68]), our choice of BioTop added considerable semantic information to our fertilizer ontology. Obviously, domain ontologies that are aligned to the same top-level ontology can be more easily integrated and related to each other. Thus, top-level ontologies have the advantage of securing similar design standards across ontology projects.

*4.4. Step 4: Identification of membership conditions*

*Purpose*

The unique advantage of logic-based ontology languages like OWL is that they allow specifying the meaning of a class through *necessary membership conditions*. The goal in this step is to identify as complete as possible characteristics that act as membership conditions. This is important because the characteristics are valuable for checking the consistency of the is-a hierarchy and infering class subsumptions automatically. It is also desirable to identify *necessary and (jointly) sufficient* membership conditions that *define* a class. The reason is that it is only defined classes under which other classes can be automatically subsumed. Nevertheless, wrongly stated membership conditions may result in the erroneous exclusion of real-life entities and/or invalid reasoning. Membership conditions serve as clear decision criteria for the membership of individuals (instances of classes) and can only be answered through yes-or-no questions.

In order to clarify the meaning of the classes, we suggest beginning with an informal (natural language) specification of the classes with membership conditions. It is not only the basis for the formal specification of the membership conditions in step 5, but also provides a stronger foundation and permits revisions of the alignments described in step 3.

*Actions to be taken*

Up to four iteratively applied actions may be necessary in this step:
  a. Identifying the meaning of the classes
  b. Collection of definitions in natural language
  c. Deciding for principles of including classes and resolving ambiguity
  d. Extraction or definition of membership conditions

(a) The primary step in the definition of membership conditions is to understand to what the concepts in the thesaurus precisely refer. For this purpose, we exploit all the means that the re-engineered thesaurus offers to express the meaning of its concepts: assigned natural language terms, hierarchy, associative relations, qualifiers, scope notes, definitions, as well as the purpose or focus of the thesaurus overall.

(b) Consulting natural language definitions from encyclopedias and dictionaries is helpful in various ways: (i) A concept may lack a definition in the thesaurus, (ii) the defining phrase of a thesaurus definition may be ambiguous or otherwise difficult to understand, (iii) there might be need for additional information to identify additional membership conditions for classes, or (iv) one wants to uncover possible ambiguities of the term to be defined. These definitions should be as domain-specific as possible in order to have a qualitatively good basis for the definition of membership conditions. Any definition needs to be in line with the meaning of a thesaurus concept. Where there are no useful definitions, it may be necessary to consult domain experts to create explicit definitions.

(c) In a thesaurus, at least one natural language term has to be assigned to a concept. For human users, these labels are the most important indicator of the meaning of a concept. Unfortunately, terms in natural language are almost always ambiguous and have different meanings in different communities and cultural contexts. Moreover, the specific understanding of a term may change over time while the term remains assigned to the same concept in the thesaurus. Sometimes terms have multiple meanings even in a single community or discipline, especially if there are dif-

ferent schools of thought. For example, one will generally have an intuitive idea of what a concept labeled "water" represents. If one asks whether a class "water" shall include instances such as water ice cubes, water in a plasma aggregate state, waste water or salt water, there may be differing opinions.

If alternative interpretations of a concept are not disambiguated by the information collected in the previous two actions, the ontology (re-)engineer is forced to make a decision between

(1) including several classes for a given term, each for every meaning,

(2) selecting one out of several possible meanings, or

(3) rejecting the inclusion of a given class entirely if it is not possible to identify any meaning that makes sense in the given context.

There are no objective criteria for such a decision. Decisions will rather depend on factors like a specific interest in using the ontology, a specific viewpoint of the ontology (re-)engineer or the relevant peer group. We recommend that such decisions be consistent and guided by transparent principles. Such principles could be, e.g., specifying material entities from a chemical point of view only or always including several classes in the case of conflicting definitions in the subject field.

(d) There exists little practical guidance for deciding whether or not (i) a membership condition is a valid *necessary* membership condition and (ii) whether one or more membership conditions constitute a set of *jointly sufficient* membership conditions for a given class. For many natural kinds of entities such as tigers or zebras, the identification of necessary and sufficient membership conditions is problematic and only necessary conditions can be indicated [69,70,34]. The specification of membership conditions may also require setting limits to decide about the membership for borderline cases. For example, one may determine a minimum amount of calcium that a calcium fertilizer needs to contain. A given material that misses the minimal amount of calcium is then not considered a calcium fertilizer, even if it misses the minimum amount just slightly.

Sometimes the hierarchical whole-part relationships or the associative relationships in the thesaurus can be adopted as membership conditions. There may, however, be kinds of entities for which no formal membership condition can be stated. In such cases, definitions should be provided in natural language or First Order Logic, which are more expressive than OWL. If need be, examples or typical characteristics can be included as comments instead of a definition.

Natural language definitions and other comments are, in any case, helpful for both ontology maintainer and user, even though they are not visible for automated reasoning programs.

*Application of the step to the fertilizer ontology*

(a) We initially attempted to understand the meaning of the fertilizer concepts in the AGROVOC thesaurus. While there are natural language terms (with or without qualifiers) and hierarchical and associative relationships for all of the concepts in AGROVOC, there are no definitions and just a few scope notes (which in AGROVOC have the character of definitions, but are rarely provided). This turned out to be a major issue for grasping the precise meaning of a concept and strongly impeded the extraction of membership conditions.

(b) We compensated for the lack of definitions in the AGROVOC thesaurus by encyclopedic and regulatory definitions. More specifically, we obtained definitions primarily from *The Fertilizer Encyclopedia* [71] and a fertilizer-related regulation by the European Commission [72]. While they covered most fertilizer classes, we sometimes had to use definitions from other sources or create customized definitions using the advice of domain experts.

(c) The definitions gathered in the previous action did not suffice to disambiguate the meaning of all concepts. Because we did neither have a specific application of the ontology in mind nor a sponsor with a specific interest, we had to decide on principles which we could follow. Since the majority of definitions adopted from *The Fertilizer Encyclopedia* related to chemical entities, we decided to stick to a scientific (chemistry-based) approach to define membership conditions wherever possible. Where there were conflicting definitions, we adopted only the commonalities of these definitions to specify necessary membership conditions.

(d) The collected definitions allowed us to grasp the meaning of concepts more precisely and to extract membership conditions. We will discuss this in detail for the concept 'fertilizer' before summarizing our work for specific fertilizer types and concepts closely related to fertilizers.

**Fertilizer**

Table 1 shows all the available information in the AGROVOC thesaurus as well as the definitions and further relevant explanatory fragments in (1) *The Fertilizer Encyclopedia* and (2) the fertilizer-related regulation of the European Commission (EC) on the

Table 1. Information revealing the meaning of 'fertilizer' in the AGROVOC thesaurus

| | |
|---|---|
| *Preferred term in the AGROVOC thesaurus* | Fertilizers |
| *Non-preferred terms in the AGROVOC thesaurus* | Fertilisers |
| *Hierarchical context in the AGROVOC thesaurus* | Fertilizers → Farm inputs → Inputs → Resources |
| *Associated concepts in the AGROVOC thesaurus (their preferred term)* | pollutants, Seed pelleting, soil amendments, Soil pollution, Balanced fertilization, Fertilizer application, Fertilizer injury, Agrochemicals, Biofertilizers, Fertilizer technology, Fertilizer industry, Foliar application, Slags, Basic slag |
| *Definition in The Fertilizer Encyclopedia* [71] | *Fertilizer:* any natural or manufactured solid or liquid material, added to the soil to supply one or more nutrients essential for the proper development and growth of a plant […] in the broadest sense, products that improve the levels of the available plant nutrients and/or the chemical and physical properties of the soil, thereby directly or indirectly enhancing the growth, yield and quality of the plant |
| *Definition in fertilizer-related regulation by the European Commission* [72] | *Fertiliser:* material, the main function of which is to provide nutrients for plants. |

concept 'fertilizer'. These form the basis for our analysis. The hierarchical context of 'fertilizer' in the AGROVOC thesaurus and a dictionary definition of 'resource' [73] suggest that fertilizer is understood as an input to farming in the AGROVOC thesaurus, farming being a kind of value production. This combines well with the definitions in *The Fertilizer Encyclopedia* and the fertilizer-related regulation by the European Commission.

The encyclopedia definition as well as the definition by the EC point to three conditions:

a)   being a material
b)   being involvable in (chemical) processes improving the plant nutrient level of soils
c)   containing nutrients for plants.

With condition (a) we summarized the description "natural or manufactured material" in the encyclopedia definition. We disregarded the limitation to "a solid *or* liquid material", as it is, in fact, not adequate. There are, for example, liquid gas fertilizers that are sold and stored as liquids, but applied in gaseous state.

The condition (b), as it is formulated, is not sufficient. There are fertilizers that are put directly onto plants, more specifically onto those parts of a plant that are not underground (i.e. not on the roots), so that the nutrients do not have to use the chemical reaction path via the soil. For this reason, we reformulated the condition (b) to express what fertilizers have to be capable of:

b*)   being able to release plant nutrients

We acknowledge that this condition may have to be further detailed, e.g., by a property of 'being water soluble' in the case of fertilizers applied on soils and a property of 'being liquid' in the case of fertilizers applied on plant leaves. This requires further detailed investigation, which we did not pursue.

The formulation of condition (c) is not satisfactory either. It is not enough for a material to contain *some* plant nutrients to be effective, but to contain *significant* amounts of plant nutrients that can actually have a fertilizing effect. Further, it is important to put the amount of plant nutrients in relation to the overall volume or mass of the fertilizer material. This modifies condition (c) as follows:

c*)   containing a significant mass proportion of plant nutrients

A more precise way of expressing the modifier "significant" is to indicate a minimum amount of plant nutrients per weight unit. For this purpose, we analyzed the fertilizer-related regulation of the European Commission [72] and the official regulation in Germany, the *Düngemittelverordnung* [74], for the fertilizer type with the lowest mass proportion of plant nutrients and adopted the mass proportion for not only 'fertilizer', but also 'compound fertilizer' and 'micronutrient fertilizer'. This turned out to be a complex study in itself that we do not further detail here. The result of our analysis was that specific kinds of micronutrient fertilizers are the types of fertilizers containing the lowest proportions of plant nutrients (plant micronutrients): a minimal mass proportion of 0.168 %. We can adopt this minimum requirement as a necessary condition for a fertilizer:

c**)   containing a minimal mass proportion of 0.168 % plant nutrients

Note, however, that these conditions are not jointly sufficient for being a fertilizer because conditions (a)–(c**) are true for many water-soluble substances with little amounts of any plant nutrient (e.g., nitrogen) that would not be considered fertilizers, e.g., various medicaments. For this reason, we will characterize fertilizers with necessary conditions only.

**Specific fertilizer types**

In the same way we analyzed 'fertilizer' in general, we also analyzed the meaning of the other fertilizer concepts and the membership conditions of the re-

spective classes. All of them have one fundamental characteristic expressed through the is-a hierarchy—being a fertilizer—and, thus, inherit all membership conditions from 'fertilizer'.

We faced similar problems to those we encountered with the class 'fertilizer' when identifying membership conditions for the classes 'compound fertilizer' and 'micronutrient fertilizer'. Compound fertilizers need to contain a minimum mass proportion of 0.27% of two or more different *primary* plant nutrients (nitrogen, sulphur or potassium). Micronutrient fertilizers need to contain at least 0.17% of plant micronutrients.

Fertilizer classes characterized by specific nutrients such as 'calcium fertilizer' or 'nitrogen phosphorus fertilizer' had the same pattern in terms of their analysis and generally refer to two membership conditions: containing a minimum mass proportion of the characterizing chemical element or molecule (e.g., 14.30% calcium or 4.50% nitrogen). These fertilizer types we could specify with necessary and sufficient conditions. An exception are the classes 'ammonium fertilizer', 'nitrate fertilizer', 'rock phosphate', 'superphosphate' and 'nitrophosphate'. We could specify them with necessary conditions only because we lacked sources that indicate minimum mass proportions of molecules by which these fertilizer types are characterized.

There are at least two different interpretations for the term 'organic fertilizer'. For the 'social' interpretation, the term refers to naturally occurring or naturally derived fertilizers; for the 'scientific' interpretation, it refers to fertilizers containing a significant mass proportion of the chemical element carbon. The social and the scientific interpretation are incompatible in the sense that they do not have the same extension in reality: unprocessed, naturally occurring mineral materials such as rock phosphate do not contain carbon—or if they do, then only in irrelevant amounts that are not type-defining. Since AGROVOC did not provide any disambiguating hint, we followed the previously established principle to use the scientific interpretation. However, we were not able to specify the carbon amount necessary for an organic fertilizer more precisely.

Specific subtypes of organic fertilizers ('biofertilizer', 'compost', 'fish manure', 'green manure' and 'guano') are generally characterized as the outcomes of specific processes with specific inputs. For example, fish manures are fish carcasses or parts of fish (offal) that have undergone the process of drying and crushing or powdering. Biofertilizers have the peculiar feature that in the very moment they are sold they are not fertilizers in the strict sense because biofertilizers are active microorganisms, bacteria or fungi that develop a symbiotic relationship with plants. At that time, they do not contain the relevant amount of plant nutrients, which conflicts with our membership conditions for the class 'fertilizer'. It is only in the course of active processes that biofertilizers *release* plant nutrients—besides having various other benefits for agriculture. It is, thus, only the material released by these organisms that can strictly be considered a fertilizer. It also remains unclear what distinguishes the plants referred to as "green manures" from other plants. Again, only the outcome of their decomposition through organisms can be considered a fertilizer, not the plant itself.

The class 'inorganic fertilizer' could only be defined as not being an organic fertilizer, which negates the containment of carbon. Organomineral fertilizers contain significant mass proportions of organic fertilizers and inorganic fertilizers; again, it was not possible to state proportions precisely enough to specify necessary and sufficient membership conditions. Liquid fertilizers and liquid gas fertilizers are characterized by the specific aggregate state in which they are applied. Slow-release fertilizers are characterized by their disposition to release plant nutrients slowly, but there is no explicit and clear-cut maximal velocity for this. Fertilizer-pesticide combinations also contain significant amounts of pesticides.

Some of the concepts in the fertilizer branch have not been included as subclasses of 'fertilizer' in the ontology, namely 'potting compost' and 'fertilizer combination'. Potting composts do not necessarily contain significant amounts of plant nutrients and fertilizer combinations are hard to be delineated precisely from other materials.

The tables in Appendix 4 provide a concise summary of all fertilizer-related classes, their membership conditions, as well as an indicator of whether the conditions are necessary ones only or whether they are also sufficient conditions expressing a definition for the class. It also contains further information about the results of the alignment process discussed in the next step.

*Discussion*

The identification of membership conditions, which underlies all subsequent steps, turned out to be the most difficult step in the re-engineering process. The first observation from our case study is that natural language definitions can facilitate the identification of necessary membership conditions. The second

observation is that identifying membership conditions stimulates thinking about the precise meaning of thesaurus concepts and whether the class hierarchy of generic relationship that was adopted from the thesaurus is free of contradictions and consistently restricts the intension.

Some of the terms we dealt with can, in natural language, refer to things in quite different states. For example, "compost" may refer to the stuff in compost piles before and after its degradation through microorganisms and "biofertilizer" may refer to either (a) a product containing organisms or (b) the product once it has been applied to the field and has bound or solubilized plant nutrients. Only the second meaning fits our definition of 'fertilizer'. In some cases, like 'potting compost', we could not think of any way in which the real-life entities could fulfill the membership conditions to be considered a fertilizer and rejected them as subclasses of 'fertilizer'. Such issues raise the question whether we have to improve the membership conditions that we specified for 'fertilizer' or other subordinate classes. They also challenge modeling decisions that have to be made between conflicting definitions. For example, we had to choose between different interpretations of 'organic fertilizer' and limit the membership conditions of the class 'plant micronutrient' to the commonalities that we found in partially conflicting definitions. Overall, the identification of membership conditions confronts one with eventual ambiguities in the thesaurus or in language in general.

Sometimes it was also difficult to decide whether a given set of necessary membership conditions is sufficient to define a class. Decisions in this respect have consequences for the reasoning results. Reviewing the inferred class hierarchy as the outcome of the reasoning included in step 5 made us revise and rethink our membership conditions. For example, we wondered whether 'compost' is the outcome of instances of the same type of decomposition as 'guano' or 'green manure' or not. It exceeded the possibility of our study to research this question further and we assumed that there is a general decomposition process or a group of such processes.

While the collection of natural language definitions from existing sources can be pursued quite mechanically, one may end up with incoherent or conflicting results. For this and other reasons, precisely specifying the frequently encountered membership condition of containing "significant amounts" of certain plant nutrients turned out to be a complex endeavor. Therefore, membership conditions cannot be considered a "nice to have" feature of an ontology.

Instead, the richness of membership conditions must be acknowledged as a key characteristic that describes the quality of an ontology and the intellectual effort that has been invested in the development of an ontology. In our case, the identification of membership conditions was tremendously time-intensive.

### 4.5. Step 5: Formal specification of membership conditions

#### Purpose

Step 5 aims to express all membership conditions gathered in the previous step in the formal language OWL using a common set of formally well-defined relationships. For this purpose, relationships as well as domain-specific ontologies are adopted and eventually amended. Because eventually-adopted domain-specific ontologies are aligned to the same top-level ontology chosen in step 3, the end result of this step is a state where the emerging ontology, the top-level ontology as well as other adopted and eventually-amended domain-specific ontologies are integrated, that is, they are commonly densely interlinked through membership conditions. Based on the formal specification of the membership conditions, an automated reasoning program (or, for short, a reasoner) can interpret and check them in combination with the class hierarchy.

#### Actions to be taken

The formal specification of classes/membership conditions can be subdivided into the following activities that are likely to be applied iteratively:
  a.  Choice of formal relations
  b.  Choice of existing domain-specific ontologies to be re-used (and, if necessary, their alignment to the top-level ontology)
  c.  "Amendment" of the external ontologies
  d.  Formalization of the class specifications
  e.  Adding natural language definitions and comments as class annotations
  f.  Consistency check and plausibility check of the fully inferred class hierarchy

(a) A fixed set of formally defined relationships (object properties in OWL) should be adopted, such as the Relation Ontology [68] or the relationships defined in BioTop [64]. The relationships are a necessary component to formally express the membership conditions that we collected in step 4. This avoids making mistakes in defining semantically precise and consistent relationships, but also enables the integration of ontologies. The adopted relationships

should have a strong tie with the adopted top-level ontology because many relationships are, and should be, constrained in their domain and range with reference to a top-level ontology. Which relationships are necessitated depends on the domain at stake, but a useful set of formally defined relationships in ontologies will generally comprise spatial, mereological and temporal relations. Most fundamental is the subclass-of (is-a) relation, which is a pre-defined part of OWL.

(b) Classes from the top-level ontology (chosen in step 3) will be useful to state very general membership conditions. For example, we will want to state that portions of agricultural fertilizers are *material objects*. Of course, the top-level ontology will not contain domain-specific classes that need to be mentioned in membership conditions. One way to supplement the top-level ontology is by adopting (i.e. reusing) existing ontologies (in part or as a whole) that cover related domains. For the biomedical field, such ontologies can be found in repositories like the Open Biomedical and Biological Ontology (OBO) Foundry [75] or via the search function of Ontobee [76] or BioPortal [77]. There are also efforts to build up ontology registries [78] and develop metadata schemes for such registries [79].

Terms can be imported from several ontologies, eventually using the support of tools like OntoFox [80] or MIREOT [81]. When one term is covered by several domain ontologies, the ontology that fits best to our method should be preferred. Particular advantages are formal membership conditions, alignment to the same top-level ontology and shared types of relationships. Otherwise, the domain ontologies that are adopted in this step need to be aligned to the same top-level ontology that was adopted in step 3–in the same way the classes derived from the thesaurus have been aligned to it.

(c) If needed classes and relationships are not found in existing ontologies, there are two options: either ontology developers can suggest these classes to be added to the respective domain ontologies or the classes are created within the emerging ontology itself. Newly created classes should, of course, not duplicate what is already contained in the imported ontologies. The introduction of new classes is, of course, unavoidable if a new domain is to be described. However, introducing new types of relationships should be avoided and considered a last resort because idiosyncratic relationships are a major obstacle for interoperability. Proliferating relationships in OWL can also severely impede the performance of the reasoning algorithms. In many cases, however, the urge to introduce new relationships is due to an

insufficient ontological analysis. E.g., the relationship 'digests' need not be introduced as a new basic type if there are already generic relations like 'agent of' and 'patient of', by means of which 'digest' can be expressed as

'agent_of' some 'digestion process' and 'has_patient'.

Newly introduced classes also need to find a place in the class hierarchy. They should be subsumed under a class in the top-level ontology or under a class in one of the (aligned) domain-specific ontologies. The assignment should be done with care because the aligned class will inherit all membership conditions of the superordinate classes. In cases of doubt, the class in question should be subsumed under a more general class.

When introducing new classes, it would, on the one hand, be desirable to fully specify them with membership conditions to enhance consistency checking and infer implicit class subsumptions. On the other hand, this is as time-consuming as re-engineering the thesaurus concepts. Moreover, the membership conditions for new classes will, in turn, refer to still other classes and so forth. We, therefore, recommend only specifying membership conditions for classes that are at the heart of the modeled domain and leaving fringe classes to specialists in these other domains. Nevertheless, in an ideal world, the membership conditions of all classes both within a single ontology and across different ontologies form a complex and interdependent network.

(d) The formal specification of classes is realized by adding the necessary membership conditions identified in step 4 as so-called anonymous superclasses using the subclass axiom. If only necessary membership conditions are stated, a class is called a *primitive class* [82]. If both necessary and sufficient conditions are given by adding them as an anonymous equivalent class, a class is called a *defined class* [82].

When formalizing membership conditions, one has to respect the formal properties of the relationships, such as their domain, range, transitivity, disjointness, inverse implication or reflexivity [43]. In cases where hierarchical whole-part relationships or associative relationships from the thesaurus have been adopted into the ontology as membership conditions, they will normally have to be refined at this stage to be matched to semantically precise formal relationships.

The formal specification of classes by membership conditions is also the step where guidelines for the correct and complete use of OWL [15,82–84] or ontology design patterns (ODPs) for standardized ontology engineering or circumventing expressivity

problems of OWL [85] should be applied. Following the guidelines may also imply defining additional axioms such as the disjointness of classes or the transitivity of relationships.

(e) Natural language definitions should be added at least when no formal specification is possible. Comments may, e.g., detail membership conditions that could not be formalized. It is important to bear in mind that information in natural language cannot be processed by automated reasoning programs.

(f) Automated reasoning allows checking the consistency of the emerging ontology on a regular basis during re-engineering. A reasoner can automatically infer new subsumptions, equivalences or other axioms if they are entailed logically by the explicitly asserted ontology.

The formalized membership conditions and the densely linked ontologies resulting from the previous activities in step 5 lay the ground based on which a reasoner can effectively check the consistency of the emerging ontology. Using a reasoner that is appropriate for the chosen OWL-DL semantics, possible contradictions can be tracked and removed.

Using a reasoner also allows distinguishing between the asserted ontology and the inferred ontology. The asserted ontology comprises the explicitly asserted statements only, while the inferred ontology also comprises all inferred statements. When speaking about "the" ontology, the reference is generally to the asserted ontology in this paper.

*Application of the step to the fertilizer ontology*

(a) Our need for formally defined relationships was entirely satisfied with the relationships contained in BioTop [65], the upper-domain ontology adopted in step 3. This was fortunate because the membership conditions in BioTop are formally specified and refer to the classes defined in BioTop. A reasoner can, thus, use the relationship specifications to check the consistency of our emerging ontology.

(b) Since our formal specifications frequently refer to chemical entities, we adopted ChEBI (= Chemical Entities of Biological Interest) [86], an ontology from the chemistry domain, the major feature of which is the completeness and hierarchical organization of the chemical elements, molecules and other entities that it models. While it was helpful that ChEBI is available in OWL format, a disadvantage of ChEBI was that it does not give explicit membership conditions (as of March 2012).

Since the range of molecules is enormous, ChEBI is a very large and complex ontology. In order to keep our ontology tractable for the automated reasoner, we extracted a fragment of less than 10% of ChEBI's original size that contains the chemical entities that are relevant for us. The slimming down was challenging, since ChEBI makes intensive use of multi-hierarchies and there was a high risk of (unintentionally) deleting branches that were to be retained because they are connected with other relevant paths at a lower level. Nevertheless, this may be considered a weakness of our chosen ontology editor, Protégé, because classes from a hierarchical path should not be deleted without user interventions if they also belong to other hierarchical paths.

Though we may have been able to adopt more ontologies, we did not search for other ontologies that could have been suitable. Searching and evaluating ontologies can be a time-consuming effort. Because our main interest was to illustrate the process of choosing and aligning external ontologies, we limited ourselves to ChEBI.

As BioTop already provides bridge classes for this purpose [64], it was not difficult to align ChEBI classes to BioTop, our chosen top-level ontology. The first three entries in table 2 show the classes that were aligned (implicitly aligning the subordinate classes) indicating the alignment axioms in the second column. We also amended some membership conditions for specific classes in ChEBI. The amended classes are listed in the last three rows of table 2. The respective entries in the 2nd column indicate the newly-asserted membership conditions.

Table 2. Amendments to imported ChEBI classes

| ChEBI Class | Amended alignment axiom or necessary membership condition |
|---|---|
| chemical entity | subclass of 'material object' (BioTop) |
| Atom | equivalent to 'atom' (BioTop) |
| Mixture | subclass of 'collective material entity' (BioTop) |
| phosphate mineral | having some 'phosphorus molecular entity' (ChEBI) as granular part |
| Calcium bis(dihydrogen-phosphate) | subclass of 'phosphorus molecular entity' (ChEBI) |
| Calcium sulfate | subclass of 'sulfur molecular entity' (ChEBI) |

(c) The specifications of the various fertilizer types required the introduction of the classes listed in table 3. Only the classes that are central to the fertilizer domain were specified with membership conditions (as done for fertilizers in step 4). A first group of these classes are the ones listed in the first row of entries in table 3 ('plant nutrient', etc.). The members of these classes are characterized by their ability to be

Table 3. Newly added classes

| Classes | Source of lowest imported superclass | Membership conditions defined? |
|---|---|---|
| plant nutrient, primary plant nutrient, secondary plant nutrient, plant micronutrient | ChEBI | Defined |
| plant nutrient disposition, plant nutrient uptake process, plant nutrient release disposition, plant nutrient release process, plant nutrient slow release disposition | BioTop | Defined |
| seabird, goat, bat, whale, portion of heterogenous gas, pesticide, binding, decomposition, solubilizing, crushing, drying, powdering, excretion | BioTop | Not defined |

picked up as nutrients by a plant. They differ in terms of the chemical elements they comprise and group the chemical elements by the quantity in which they are required by plant nutrients.

We also introduced classes for processes and dispositions [87]. E.g., we introduced a class 'plant nutrient uptake disposition' comprising all instances of the disposition to *be picked up* as plant nutrient, whereas the class 'plant nutrient release disposition' comprises all instances of the ability to *release* plant nutrients. The 'plant nutrient uptake process' and the 'plant nutrient release process' are the corresponding process types that realize these dispositions. Plant nutrient uptake processes take place in plants and have plants as well as plant nutrients as participants.

We did not introduce new relationships in the current development step because BioTop already contained all relationships needed—with the exception of some relationships that were tentatively used in subactivity (d) to represent certain features as data properties.

Table 3 indicates the ontology (BioTop or ChEBI) to which the newly introduced classes have been aligned. While not listing the precise alignments here, we always chose the most specific class in the ontology to which we aligned. Nevertheless, we only stated alignments about which we were confident; for this reason, we sometimes aligned to quite general classes.

(d) The natural language formulations of the membership conditions are concisely summarized in Appendix 4. They are already formulated with an eye on the relationships provided by BioTop, so that they translate well into OWL class expressions. Only some classes like 'fish manure' and 'guano' have complex membership conditions and, thus, also com-

plex formal expressions. The phrase 'being a', as used in the natural language formulations of membership conditions in previous steps, translates into the OWL axiom 'SubClassOf'. In the case of classes that are defined with necessary and sufficient conditions, the 'EquivalentTo equivalentTo' axiom is used; the subclass conditions become part of the class expression that is asserted to be equivalent.

The formal specification of the membership conditions to contain a minimal proportion of plant nutrients turned out to be problematic because the expressivity of OWL2 does not provide a straightforward way to express proportions. Simply adding annotations is easy to implement, but the quantification is not machine-readable then. Using the minimum modifier for a relationship (the ObjectMinCardinality axiom), e.g.,

'has granular part' *min* 1680 'plant nutrient',

has the advantage that the restriction is explicit and machine-readable. Unfortunately, the minimum qualifier for object properties in OWL does not express proportions, but rather countable quantities. In consequence, the condition stated above expresses that fertilizer must contain at least 1680 individually countable plant nutrients. This problem is also not addressed by creating a subtype of the 'has granular part' relationship that expresses in its label the desired semantics, e.g.,

'contains mass proportion (in ppm) of granular part' *min* 1680 'plant nutrient'.

Automated reasoning algorithms cannot recognize the intended semantic difference in the relationship label and would still interpret the modifier as a condition in a countable sense. As a workaround, we tentatively resorted to Data properties in OWL (the DataMinCardinality axiom), e.g.,

'contains nutrient mass proportion of (in ppm)' min 1680 integer

which are not pre-occupied with what their values express. Nevertheless, this condition needs the company of the condition

'has granular part' *some* 'plant nutrient'

rather than substituting it. This solution provides machine-readability of the quantity, but it necessitates a hierarchy of data properties that parallels the hierarchy of chemical elements having the disposition to act as a plant nutrient. For example, in order to express the containment of calcium in 'calcium fertilizer' there has to be created a data property 'contains
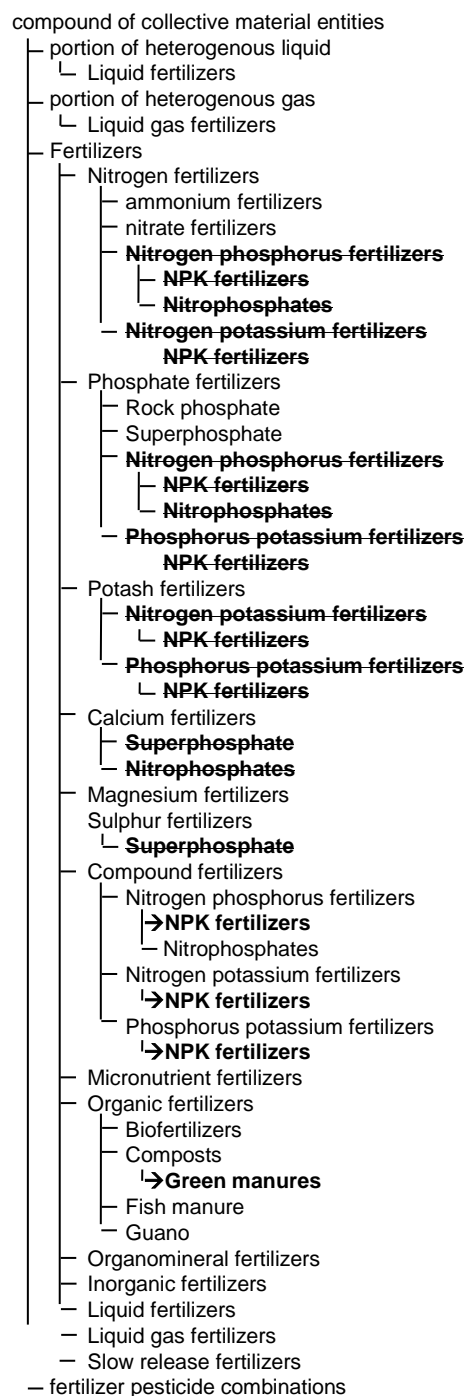
*calcium* mass proportion of (in ppm)' subordinated to 'contains *nutrient* mass proportion of (in ppm)'. Despite not being very elegant, we chose to apply this solution to address the expressivity problem of OWL. Another strategy to express minimum quantities is to supplement the ontology proper with additional tools like a database, but this is not part of ontological modeling in the strict sense and, thus, not in the scope of the present case study.

A similar problem for working with relationships in OWL and Protégé is that they cannot be combined with quantities in percentages, but only with cardinalities given in natural numbers. This problem can be solved by scaling the values and expressing them as parts per million (abbreviated ppm) with respect to the mass proportion as was done in the examples above. The minimum plant nutrient proportions in percentages were transferred into a parts per million (ppm) measure, i.e. a value of 1680 refers to a share of 1680 millionths of the number of particles (=0.168%). All measures, including the ppm measure, refer to mass proportions (as opposed to a volume proportion).

(e) We refrained from adding the natural language definitions identified in step 4 as class annotations because we often modified these definitions with the goal of expressing more precise membership conditions.

(f) Based on the formal specification of the aligned ontology with its membership conditions for the various classes, we were able to check the ontology for consistency in a non-trivial way and infer subsumptions in the class hierarchy that have not already been asserted. For this purpose we used the reasoner Hermit [88], which is available as an embedded plug-in for the Protégé-OWL editor.

The reasoning process revealed various initial modeling mistakes that we subsequently resolved. Many mistakes were technical ones, similar to the typical mistakes described in guidelines [15,82,83]. Other mistakes revealed an insufficient understanding of the adopted (imported) ontologies and relationships. The reasoning results also made us wonder whether we made mistakes in asserting is-a relationships and membership conditions. For example, we felt uncertain about the correctness of the inferred subsumptions of 'NPK fertilizer' and 'green manure' indicated in figure 8, but could not finally find arguments against them. In the case of the NPK fertilizers, this uneasiness is, of course, due to the ambiguity connected to superordinate terms like 'Compound fertilizers'.

compound of collective material entities
├─ portion of heterogenous liquid
│  └─ Liquid fertilizers
├─ portion of heterogenous gas
│  └─ Liquid gas fertilizers
├─ Fertilizers
│  ├─ Nitrogen fertilizers
│  │  ├─ ammonium fertilizers
│  │  ├─ nitrate fertilizers
│  │  ├─ **Nitrogen phosphorus fertilizers**
│  │  │  ├─ **NPK fertilizers**
│  │  │  └─ **Nitrophosphates**
│  │  └─ **Nitrogen potassium fertilizers**
│  │        **NPK fertilizers**
│  ├─ Phosphate fertilizers
│  │  ├─ Rock phosphate
│  │  ├─ Superphosphate
│  │  ├─ **Nitrogen phosphorus fertilizers**
│  │  │  ├─ **NPK fertilizers**
│  │  │  └─ **Nitrophosphates**
│  │  └─ **Phosphorus potassium fertilizers**
│  │        **NPK fertilizers**
│  ├─ Potash fertilizers
│  │  ├─ **Nitrogen potassium fertilizers**
│  │  │  └─ **NPK fertilizers**
│  │  └─ **Phosphorus potassium fertilizers**
│  │        └─ **NPK fertilizers**
│  ├─ Calcium fertilizers
│  │  ├─ **Superphosphate**
│  │  └─ **Nitrophosphates**
│  ├─ Magnesium fertilizers
│  │  Sulphur fertilizers
│  │  └─ **Superphosphate**
│  ├─ Compound fertilizers
│  │  ├─ Nitrogen phosphorus fertilizers
│  │  │  ├─ →**NPK fertilizers**
│  │  │  └─ Nitrophosphates
│  │  ├─ Nitrogen potassium fertilizers
│  │  │  └─ →**NPK fertilizers**
│  │  └─ Phosphorus potassium fertilizers
│  │        └─ →**NPK fertilizers**
│  ├─ Micronutrient fertilizers
│  ├─ Organic fertilizers
│  │  ├─ Biofertilizers
│  │  ├─ Composts
│  │  │  └─ →**Green manures**
│  │  ├─ Fish manure
│  │  └─ Guano
│  ├─ Organomineral fertilizers
│  ├─ Inorganic fertilizers
│  ├─ Liquid fertilizers
│  ├─ Liquid gas fertilizers
│  └─ Slow release fertilizers
└─ fertilizer pesticide combinations

**bold font**…inferred subsumption
→…class subsumed under a former sibling term
~~struck through~~…incorrectly subsumed in the absence of (temporarily removed) conditions concerning minimum proportions of the respective plant nutrient(s)

Figure 8. Inferred fertilizer class hierarchy after alignment

It also turned out that there are considerable problems with reasoning over the data properties that we introduced as described above. When defining values for the data properties that are greater than 1000, Hermit aborted the initialization of the reasoning process with error messages. Moreover, the computing time increased tremendously when using data properties in the fertilizer class definitions. While the first problem could have been avoided by indicating the mass proportions in per mill (thousandths) instead of millionths and rounding them, attempts to improve the performance by dissolving the data property hierarchy were not successful.

It is outside the scope of this paper to determine whether the problem with the data properties is a general one or a particular problem of the Hermit reasoner. In the end, the data properties had to be removed from the class specifications to be able to use the reasoner. In consequence, the concerned class specifications became primitive ones with insufficient membership conditions. This, in turn, results in the loss of desirable reasoning inferences since new class subsumptions can only be inferred under classes defined with necessary and sufficient conditions. We checked manually whether valid subsumptions can be inferred based on the minimum plant nutrient levels. For this purpose, we left out all conditions involving data properties. This would normally imply that the defined classes become primitive classes, but we kept their status as defined classes, which expectedly led to wrongly inferred subsumptions. We addressed the problem by manually sorting out wrongly inferred subsumptions, which are struck through in figure 8. After the critical review, we had left only very few (correctly inferred) new is-a relationships that were not stated in the previously asserted class hierarchy (indicated in bold font in figure 8). The NPK fertilizers were subsumed deeper in the hierarchy under nitrogen phosphorus fertilizer, nitrogen potassium fertilizer as well as under phosphorus potassium fertilizer, which is a plausible result.

The subsumption of 'green manure fertilizer' under 'compost' appears more debatable. It results from the assumption that plants used as green manure undergo the same degradation process as other material that is usually referred to as "compost".

*Discussion*

BioTop provided a comprehensive set of part-of and other relationships whose formal specifications refer to the very same top-level categories that we adopted for our alignment in step 3. The relationships satisfied our modeling needs entirely and, again, took decisions from us and potentially avoided wrong conclusions and modeling mistakes.

The problems faced with the formal specification of classes through membership conditions demonstrated clearly that the expressivity of a formal language can impede the formally correct specification of membership conditions. One can even be forced to remove membership conditions that have been identified earlier. In consequence, classes may lose their quality of being specified through necessary and sufficient membership conditions (being defined classes).

*4.6. Step 6: Adjustment of spelling, punctuation and other aspects of entity labels*

*Purpose*

In this step, a naming convention is chosen and the labels of classes and other entities are adjusted accordingly. This improves both readability and intelligibility of the ontology for ontology developers and users. Further, one can observe that the labels in ontologies are meant to express the context-free meaning (intension) of a class as precise as possible. While being highly recommended for maintenance and other possible usage reasons, the labeling does not change the semantics of a class for computers.

*Actions to be taken*

The adjustment involves two sequential steps:

a. Choice of a labeling convention
b. Adjusting the class labels

Currently, there are no universally accepted conventions on how ontology classes should be labeled [89]. Nevertheless, common practices have been summarized [90] and it ought to be checked whether similar conventions exist in one's field. For example, it appears to be generally accepted that names for ontology classes should be in their singular form. In any case, care should be taken to apply a single naming style consistently for all classes of the ontology.

It should be noted that the labeling described here does not concern the unique identifier (URI/IRI) of the classes or properties as specified in RFC 3986 [49]. Synonym sets from the source thesaurus could be transferred to the ontology using the labeling provisions of the respective ontology language, but this does not belong to the ontological content in the strict sense, though the integration of synonyms may be useful for some applications of the ontology.

*Application of the step to the fertilizer ontology*

We adopted common conventions in biomedical ontologies for the class labeling summarized by Schober et al. [90]. The application of the conventions often changed the first letters from upper case to lower case and also the plural forms, which are often used in thesauri, have been changed into the singular form of the nouns. The abbreviation 'NPK' (standing for nitrogen, phosphorus and potassium) is an exception and we left it unchanged because lower case letters would make the class label confusing. For example, the thesaurus concept with the preferred term 'Fertilizers' was labeled 'fertilizer' when modeled as a class in the ontology.

The identified membership conditions motivated us to change the formulations of some class labels. All fertilizer types were re-labeled to begin with "portion of" to emphasize that we deal with amounts of materials, not with countable objects. The word "fertilizer" was added to the class labels "rock phosphate", "superphosphate" and "nitrophosphate" to indicate their use as fertilizers. The ending "fertilizer" was also added to the labels of various subclasses of the 'organic fertilizer' class: 'compost', 'fish manure', 'green manure' and 'guano'. In these cases, the ending "fertilizer" emphasizes that it is not the bare organic material put on a compost heap, the unprocessed fish manure, the plant biomass called 'green manure', or the excrements of certain animals themselves that act as the fertilizer, but rather only the outcome of specific processes to which the previously mentioned materials are input. In the case of 'fish manure', we adopted the commonly used term "fish fertilizer". Appendix 5 provides a complete overview of the labeling changes.

### 4.7. Step 7: Dissolving poly-hierarchies

*Purpose*

In order to achieve an ontology that can easily be maintained, an asserted ontology should consist of a single is-a hierarchy. Thus, poly-hierarchies should be dissolved. Recall that we removed semantically incorrect poly-hierarchies that lead to contradictory membership conditions already in step 3. Dissolving semantically correct poly-hierarchies is an optional step insofar as it changes the structure but not the semantics of the ontology. Mono-hierarchies are easier to implement and maintain, but sometimes it might be intellectually challenging to decide which is-a relation is to be dissolved.

*Actions to be taken*

Assume that a target class X has two or more direct superclasses. In order to dissolve the poly-hierarchy, we have to solve two problems: Which class is to be kept as a direct superclass? And how do we retain the information given by the subclass-of relationship connecting X to the other superclasses? A clear case for discarding direct superclasses is when they are specified with necessary and sufficient membership conditions that can also be directly applied to X. In general, the strategy for retaining the information is to (a) add the restrictions of the classes along the dissolved class paths to the specification of the target class X and (b) remove any subsumption of the target class under classes of the dissolved class path from the specification of the target class X. This way, the poly-hierarchy is dissolved in the asserted ontology, but it will be restored by generating the inferred ontology through automated reasoning. Dissolving poly-hierarchies in such a way is one aspect of the *normalization method* recommended by Rector [61].

*Application of the step to the fertilizer ontology*

In the ontology that we have modeled, there are only two classes that are poly-hierarchically subsumed under several classes: 'liquid fertilizer' and 'liquid gas fertilizer'. Since dissolving the poly-hierarchy is to be handled in the same way in these two cases, we will only discuss the poly-hierarchy of the class 'portion of liquid fertilizer' here, illustrated in figure 9.
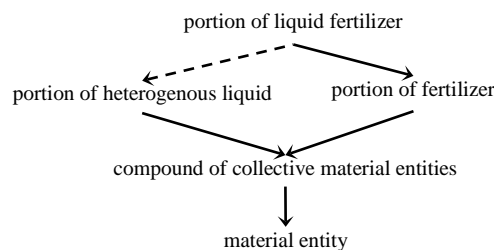


Figure 9. Poly-hierarchy for 'liquid fertilizer' (the dotted arrow indicates the is-a relationship dissolved by us).

We decided to resolve the poly-hierarchy by making 'liquid fertilizer' primarily belong to the class 'fertilizer'. Thus, we replaced the hierarchical subsumption under 'portion of heterogenous liquid' (indicated through a dotted arrow in figure 9) by adding a membership condition to the specification of the class 'liquid fertilizer', namely

'bearer of' some ('quality located' some
'liquid value region')

Of course, membership conditions that are already part of the 'liquid fertilizer' specification or its super-classes along the retained class path do not have to be added again to the specification. The formal specification of the class changes as follows:

*Before dissolving poly-hierarchy:*
'portion of liquid fertilizer' EquivalentTo
    (fertilizer and 'portion of heterogenous liquid')

*After dissolving poly-hierarchy:*
'portion of liquid fertilizer' EquivalentTo
    (fertilizer and ('bearer of' some
        ('quality located' some 'liquid value region')))

The subsumption under 'portion of heterogenous liquid' will be restored finally in the inferred class hierarchy.

## 5. Overall discussion of the re-engineering method

In the previous section, we have discussed the various steps of our re-engineering method. They are concisely summarized in Appendix 2, including all sub-activities. In this section, we will reflect on the method as a whole, in particular the benefit and effort of applying it, its generality and limitations.

The overarching motivation for the steps in our method was to re-engineer thesauri into semantically adequate ontologies that (a) make full and correct use of the semantic expressivity of OWL, (b) facilitate the integration of the ontologies with other ontologies following the same development principles, and (c) are consistent and provide reasoning results that correspond to the represented domain. Our method achieves this goal by addressing the following requirements:

(1) The ontology is described in a well-defined syntax and adheres to the description logic semantics (steps 2 and 5).
(2) The meaning of classes is expressed by means of membership conditions (steps 4 and 5).
(3) Newly created as well as imported classes are aligned to a top-level ontology; a common set of formal relationships is used (steps 3 and 5).
(4) The ontology is logically consistent and inferences from the asserted ontology are plausible (step 5).
(5) The ontology has a rigorous is-a hierarchy in which the intension of classes (the specification of the classes) becomes more restrictive at every subordinate level (steps 3-5).

(6) Natural language terms either reflect the meaning of a class as precisely as possible *or* the membership conditions of a class are intended to define one understanding of a natural language term.

Requirement (5) is probably the least obvious; it is put into effect by the adoption of the generic relationships in a thesaurus as an is-a hierarchy and its gradual refinement by basing it on membership conditions (step 5), adopting high-level membership conditions through the alignment to a top-level ontology (step 3) and, finally, checking the is-a hierarchy for its consistency (step 5).

The overall benefits of a semantically adequate ontology, as opposed to a thesaurus, need to be further investigated. The rigorous is-a hierarchy makes ontologies especially suited for automated processing, like automatic classifications and clustering. Another particular usage of an ontology is to assure interoperability among databases. Moreover, it might also be easier to maintain an ontology than a thesaurus. The comparative performance of thesauri and ontologies in natural language processing or information retrieval may depend on the specific application scenario. Because of the many structural changes and the removal of many relationships from a thesaurus, an ontology cannot always be assumed to be better than a thesaurus.

The effort of applying our re-engineering method was considerable. By far the biggest effort lies in specifying the intension of the respective concepts/classes with necessary and possibly sufficient membership conditions (steps 4 and 5). Determining minimum proportions of plant nutrients in fertilizers and formalizing these in OWL has literally become a study in its own right. It also took considerable time to become adjusted to the framework of BioTop and the ChEBI ontology to express the membership conditions using these ontologies (step 5). For example, it was not clear to us whether we should model fertilizers as having a *disposition*, a *function* or a *role* to release plant nutrients to assign them to the class 'fertilizer'. Distinguishing between these classes is, indeed, difficult and could require better clarification [91]: A role would stress the social aspects involved in producing, selling, applying and regulating certain substances as fertilizers, but then everything that would ever be sold by quacks and used by superstitious gardeners would count as a fertilizer. In contrast, functions are essential features of their bearers, being intended by designers or selected for by an evolutionary process. While this might be feasible for artificial chemical fertilizers, it does not fit for biofertilizers. It

would be strange, for example, to assume that guano has been selected in an evolutionary process for its fertilizing features. For this reason, we opted for modeling fertilizers as bearers of a disposition to release plant nutrients, in line with the overall scientific outlook of the project.

The effort of thesaurus re-engineering and ontology engineering in general can be reduced under certain circumstances:

- The effort to prepare and check the thesaurus (step 1) depends on the quality of the existing thesaurus. Ideally, it can be skipped entirely.
- The involvement of domain experts can save time during the identification of membership conditions (step 4).
- Experience with the chosen top-level ontology and imported ontologies reduces the alignment effort (step 3) as well as the effort to formalize the membership conditions (step 5).
- Experience in modeling with OWL reduces the effort to formally specify membership conditions correctly (step 5).
- Optional steps and sub-activities such as adjusting entity labels (step 6) or dissolving poly-hierarchies (step 7) may be omitted (see appendix 2 for an overview of optional steps).
- Steps 2, 6, and 7 may be at least partially automatable while the other steps appear to have no automation potential at the current state of the art without substantial quality losses.

During re-engineering, we faced various difficult decisions such as which top-level ontology to adopt and when to adopt classes from other domain-specific ontologies (as opposed to creating new classes), and which sources to choose. These decisions could be facilitated by way of standardization as suggested by the strategy of the OBO Foundry to collect orthogonal reference ontologies for well-defined domains [8]. Another problem was to decide for which of the classes we should give detailed membership conditions (as opposed to simply subsuming them under some existing classes). In practice, this question will often have to be answered with an eye on the prospective use of the ontology.

The alignment to a top-level ontology (step 3) as well as defining and formalizing membership conditions (steps 4 and 5) are the key features of our method because they express the meaning (the semantics)

of entities and improve the is-a hierarchy. For the most part, it is the application of these three steps in ontology engineering that makes OWL-DL-compliant ontologies differ from thesauri and distinguish our method from the re-engineering approaches mentioned in section 2.2. We believe that these steps are essential for a reliable integration of classes and class hierarchies on the Semantic Web.

Re-engineering a thesaurus into an ontology based on our method can change the structure of the thesaurus considerably. Not only the hierarchy of the thesaurus may be changed fundamentally, but also many relationships of the thesaurus may have to be rejected. These and further differences between thesauri and OWL-DL ontologies are described in greater detail in Kless et al. [92]. Differences and similarities between thesauri and ontologies have already been theoretically anticipated in a prior comparative study of relata and relationships in thesauri and ontologies [42]. However, we did not face all these differences in this case study. For example, there was no need to set apart generic relationships (is-a relations) from other types of hierarchical thesaurus relationships. The method describes the need to address such issues, but had no opportunity to collect practical experience during the re-engineering of the fertilizer branch.

Many of the steps that we have adopted in our re-engineering method have been successfully applied in ontology engineering in the life sciences. It is an open question whether one faces bigger problems when applying our method in other domains such as the social sciences. For example, it is not at all trivial to align phenomena like 'freedom' or 'success' to a top-level class. This does not question the applicability of our re-engineering method as such, but rather points to ontological problems that arise when dealing with such phenomena.

Our case study used a thesaurus as a starting point for re-engineering and could, thus, rest on a given number of existing concepts, terms and relationships. Nevertheless, a great part of the method is not specific to thesauri, but could be seen as a method of ontology engineering and re-engineering *in general*, in particular steps 3-7. This makes the method adaptable for the re-engineering of other types of structured vocabularies such as classification schemes. In fact, there are no specified minimum requirements for when a given vocabulary can be classified as a thesaurus. For this reason, but also because the quality of real-life thesauri differs significantly, we included an "up-grading" of the thesaurus according to ISO 25964-1:2011 as a first step in our method. Whether such up-grading is worth the effort must be judged in

each case separately. Alternatively, the ontology could be developed from scratch or the relationships in the thesaurus could be neglected during the re-engineering process.

## 6. Conclusions

We presented a method with seven steps and numerous sub-activities for re-engineering thesauri into semantically adequate ontologies using the description-logic-based Web Ontology Language (OWL). We motivated each step in our method and gave a detailed explanation of the activities for its realization. Further, we demonstrated the applicability of the method by applying it to a portion of the AGROVOC thesaurus that is concerned with agricultural fertilizers.

The method is applicable to all thesauri that follow the basic structure laid out in the current ISO standard for thesauri and its predecessors. It differs from previous re-engineering approaches by making full use of OWL to specify the meaning of concepts. The major strength of this method lies in producing ontologies that are truthful representations of things in reality and can be integrated consistently with each other. These benefits are achieved by imposing a rigid is-a hierarchy and removing relationships and other content from thesauri which are not appropriate for a formal ontology.

## 7. Acknowledgements

## 8. References

[1]    E. Simperl, C. Tempich, Y. Sure, Ontocom: A cost estimation model for ontology engineering, in: I. Cruz, S. Decker, D. Allemang, C. Preist, D. Schwabe, P. Mika, et al. (Eds.), The Semantic Web - ISWC 2006, Springer, Berlin; Heidelberg, 2006: pp. 625–639. doi:10.1007/11926078.

[2]    E. Simperl, Reusing ontologies on the Semantic Web: A feasibility study, Data & Knowledge Engineering. 68:10 (2009) 905–925. doi:10.1016/j.datak.2009.02.002.

[3]    M. van Assem, Converting and Integrating Vocabularies for the Semantic Web, PhD thesis, Vrije Universiteit Amsterdam, 2010.

[4]    B.M. Villazón-Terrazas, A Method for Reusing and Re-engineering Non-ontological Resources for Building Ontologies, PhD thesis, Universidad Politécnica de Madrid, 2011.

[5]    F. Baader, W. Nutt, Basic Description Logics, in: F. Baader, D. Calvanese, D.L. McGuinness, D. Nardi, P.F. Patel-Schneider (Eds.), The Description Logic Handbook: Theory, Implementation, and Applications, 2nd ed., Cambridge University Press, Cambridge, 2003: pp. 43–95.

[6]    W3C OWL Working Group, OWL 2 Web Ontology Language. Document Overview (Second Edition), World Wide Web Consortium (W3C), 2012. http://www.w3.org/TR/2012/REC-owl2-overview-20121211/.

[7]    B. Motik, P.F. Patel-Schneider, B.C. Grau, eds., OWL 2 Web Ontology Language. Direct Semantics, World Wide Web Consortium (W3C), 2009. http://www.w3.org/TR/2009/REC-owl2-direct-semantics-20091027/.

[8]    OBO Foundry homepage, The Open Biological and Biomedical Ontologies, (2012). http://www.obofoundry.org/ (accessed October 25, 2012).

[9]    U. Hahn, Turning Informal Thesauri into Formal Ontologies: A Feasibility Study on Biomedical Knowledge Re-Use, Comparative and Functional Genomics. 4:1 (2003) 94–97. doi:10.1002/cfg.247.

[10]   U. Hahn, S. Schulz, Ontology engineering by thesaurus re-engineering, in: H. Kangassalo, H. Kangassalo, H. Jaakkola, E. Kawaguchi (Eds.), Information Modelling and Knowledge Bases XIII, IOS Press, Amsterdam, 2002.

[11]   C. Wroe, R. Stevens, C.A. Goble, M. Ashburner, A methodology to migrate the Gene ontology to a description logic environment using DAML OIL, in: R.B. Altma, A.K. Dunker, L. Hunter, T.A. Jung, T.E. Klein (Eds.), Proceedings of the 8th Pacific Symposium on Biocomputing (PSB), World Scientific Publishing, Singapore, 2003: pp. 624–635.

[12]   E. Cardillo, A. Folino, R. Trunfio, R. Guarasci, Towards the Reuse of Standardized Thesauri

Into Ontologies, in: V. de Boer, A. Gangemi, K. Janowicz, A. Lawrynowicz (Eds.), Proceedings of the 5th Workshop on Ontology and Semantic Web Patterns (WOP2014), Riva del Garda, Italy, October 19, 2014 (= CEUR 1302), http://ceur-ws.org/Vol-1302, pp. 26–37.

[13] S. Angeletou, H. Lewen, B. Villazón, Methods for re-engineering and evaluation, NeOn Project, Deliverable 2.2.4, http://www.neon-project.org/deliverables/, 2010.

[14] M. Sabou, M. d'Aquin, E. Motta, SCARLET: SemantiC RelAtion DiscoveRy by Harvesting OnLinE OnTologies, in: S. Bechhofer, M. Hauswirth, J. Hoffmann, M. Koubarakis (Eds.), The Semantic Web: Research and Applications, Springer, Berlin, Heidelberg, 2008: pp. 854–858.

[15] A. Borgida, R.J. Brachman, Conceptual Modeling with Description Logics, in: F. Baader, D. Calvanese, D.L. McGuinness, D. Nardi, P.F. Patel-Schneider (Eds.), The Description Logic Handbook: Theory, Implementation, and Applications, 2nd ed., Cambridge University Press, 2003: pp. 349–372.

[16] N.F. Noy, D.L. McGuinness, Ontology Development 101: A Guide to Creating Your First Ontology, Stanford University, Stanford, U.S.A., 2001.

[17] N. Guarino, C. Welty, An Overview of Onto-Clean, in: S. Staab, R. Studer (Eds.), Handbook on Ontologies, 2nd ed., Springer, Berlin; Heidelberg, 2009: pp. 201–220.

[18] R. Rocha Souza, D. Tudhope, M. Barcellos Almeida, The KOS spectra: A tentative typology of knowledge organization systems, in: C. Gnoli, F. Mazzocchi (Eds.), Paragdigms and Conceptual Systems in Knowledge Organization: Proceedings of the 11th ISKO International Conference, 23-26 February 2010, Rome, Italy, Ergon, Würzburg, 2010: pp. 122–129.

[19] ANSI/NISO Z39.19-2005, Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies, National Information Standards Organization, 2005.

[20] ISO 25964-1:2011, Information and documentation -- Thesauri and interoperability with other vocabularies -- Part 1: Thesauri for information retrieval, International Organization for Standardization, 2011.

[21] Taxonomy Warehouse, Taxonomies, (2012). http://www.taxonomywarehouse.com/headword_list_new.aspx?vObject=10076&stype=ab (accessed December 8, 2012).

[22] AGROVOC, Agricultural Information Management Standards (AIMS). (2012). http://aims.fao.org/standards/agrovoc/about (accessed November 12, 2012).

[23] D. Kless, L. Jansen, J. Lindenthal, J. Wiebensohn, A method for re-engineering a thesaurus into an ontology, in: M. Donnelly, G. Guizzardi (Eds.), Formal Ontology in Information Systems. Proceedings of the 7th International Conference (FOIS 2012), IOS Press, Amsterdam; Washington, DC, 2012: pp. 133–146. doi:10.3233/978-1-61499-084-0-133.

[24] F. Baader, I. Horrocks, U. Sattler, Description Logics, in: S. Staab, R. Studer (Eds.), Handbook on Ontologies, 2nd ed., Springer, Berlin; Heidelberg, 2009: pp. 21–43.

[25] D. Nardi, R.J. Brachman, An Introduction to Description Logics, in: F. Baader, D. Calvanese, D.L. McGuinness, D. Nardi, P.F. Patel-Schneider (Eds.), The Description Logic Handbook: Theory, Implementation, and Applications, 2nd ed., Cambridge University Press, Cambridge, 2003: pp. 1–40.

[26] M.A. Jeusfeld, Metamodel, in: L. Liu, M.T. Özsu (Eds.), Encyclopedia of Database Systems, Springer, USA, 2009: pp. 1727–1730.

[27] N. Guarino, D. Oberle, S. Staab, What is an Ontology?, in: S. Staab, R. Studer (Eds.), Handbook on Ontologies, 2nd ed., Springer, Berlin; Heidelberg, 2009: pp. 1–17.

[28] L. Jansen, S. Schulz, Formal ontologies in biomedical knowledge representation, IMIA Yearbook 2013. IMIA Yearbook of Medical Informatics (2013) 132–146.

[29] B.A. Brody, Logical terms, glossary of, in: D.M. Borchert (Ed.), Encyclopedia of Philosophy, 2nd ed., Vol. 5 of 10, Macmillan Reference, 2006: pp. 533–560.

[30] Database model, Wikipedia, the Free Encyclopedia. (n.d.). http://en.wikipedia.org/wiki/Database_model (accessed April 17, 2013).

[31] S. Schulz, D. Schober, I. Tudose, H. Stenzhorn, The Pitfalls of Thesaurus Ontologization–The Case of the NCI Thesaurus, in: AMIA Annual Symposium Proceedings, American Medical Informatics Association, 2010: p. 727.

[32] B.C. Grau, I. Horrocks, B. Motik, B. Parsia, P.F. Patel-Schneider, U. Sattler, OWL 2: The next step for OWL, Web Semantics: Science, Services and Agents on the World Wide Web. 6:4 (2008) 309–322. doi:10.1016/j.websem.2008.05.001.

[33] R. Stevens, M. Egaña Aranguren, K. Wolsten-croft, U. Sattler, N. Drummond, M. Horridge, et al., Using OWL to model biological knowledge, International Journal of Human-Computer Studies. 65:7 (2007) 583–594. doi:10.1016/j.ijhcs.2007.03.006.

[34] J.I. Saeed, Semantics, 3rd ed., Wiley-Blackwell, Chichester, UK, 2009.

[35] D. Soergel, B. Lauser, A. Liang, F. Fisseha, J. Keizer, S. Katz, Reengineering Thesauri for New Applications: the AGROVOC Example, Journal of Digital Information. 4:4 (2004).

[36] A. Kawtrakul, A. Imsombut, A. Thunkijjanukit, D. Soergel, A. Liang, M. Sini, et al., Automatic term relationship cleaning and refinement for AGROVOC, in: J.B. Cunha, R. Morais (Eds.), Proceedings of the 5th Conference of the European Federation for Information Technology in Agriculture, Food and Environment (EFITA/WCCA 2005), July 25-28, 2005 Vila Real, Portugal, Universidade de Trás-os-Montes e Alto Douro, Vila Real, Portugal, 2005: pp. 1146–1154.

[37] S. Sánchez-Alonso, M.-A. Sicilia, Using an AGROVOC-based ontology for the description of learning resources on organic agriculture, in: M.-A. Sicilia, M.D. Lytras (Eds.), Metadata and Semantics, Springer, U.S.A., 2007: pp. 481–492. doi:10.1007/978-0-387-77745-0_47.

[38] D. Tudhope, H. Alani, C. Jones, Augmenting Thesaurus Relationships: Possibilities for Retrieval, Journal of Digital Information. 1 (2001).

[39] RDF Working Group, Resource Description Framework (RDF), (2013). http://www.w3.org/RDF/.

[40] P. Hayes, RDF Semantics, World Wide Web Consortium (W3C), 2004. http://www.w3.org/TR/2004/REC-rdf-mt-20040210/ (accessed March 11, 2010).

[41] A. Isaac, Homepage, SKOS: Simple Knowledge Organization System. (2012). http://www.niso.org/schemas/iso25964/ (accessed September 17, 2012).

[42] D. Kless, S. Milton, E. Kazmierczak, Relationships and Relata in Ontologies and Thesauri: Differences and Similarities, Applied Ontology. 7:4 (2012) 401–428. doi:10.3233/AO-2012-0118.

[43] B. Motik, P.F. Patel-Schneider, B. Parsia, eds., OWL 2 Web Ontology Language. Structural Specification and Functional-Style Syntax, World Wide Web Consortium (W3C), 2009. http://www.w3.org/TR/2009/REC-owl2-syntax-20091027/.

[44] P. Hitzler, M. Krötzsch, B. Parsia, P.F. Patel-Schneider, S. Rudolph, eds., OWL 2 Web Ontology Language. Primer, World Wide Web Consortium (W3C), 2009. http://www.w3.org/TR/owl2-primer/.

[45] Protégé-OWL editor, What is Protégé-OWL?, (2012). http://protege.stanford.edu/overview/protege-owl.html (accessed October 25, 2012).

[46] K. La Barre, Facet analysis, Annual Review of Information Science and Technology. 44:1 (2010) 243–284. doi:10.1002/aris.2010.1440440113.

[47] A.L.D. Brockmöller, Ontological Thesaurus Extension: the AAT example, PhD thesis, University of Amsterdam (UvA), 2003.

[48] A. Isaac, E. Summers, eds., SKOS Simple Knowledge Organization System. Primer, World Wide Web Consortium (W3C), 2009. http://www.w3.org/TR/2009/NOTE-skos-primer-20090818/.

[49] RFC 3986, Internationalized Resource Identifiers (IRIs), 2005. http://www.ietf.org/rfc/rfc3987.txt.

[50] S. Borgo, C. Masolo, Foundational Choices in DOLCE, in: S. Staab, R. Studer (Eds.), Handbook on Ontologies, 2nd ed., Springer, Berlin; Heidelberg, 2009: pp. 361–381.

[51] A.D. Spear, Ontology for the Twenty First Century: An Introduction with Recommendations, IFOMIS, Saarbrücken, Germany, 2006.

[52] H. Herre, General Formal Ontology (GFO): A Foundational Ontology for Conceptual Modelling, in: R. Poli, M. Healy, A. Kameas (Eds.), Theory and Applications of Ontology: Computer Applications, Springer, Netherlands, Dordrecht, 2010: pp. 297–345.

[53] E. Bertino, B. Catania, G.P. Zarri, The Cyc project, in: Intelligent Database Systems, Addison-Wesley Professional, 2001: pp. 275–316.

[54] I. Niles, A. Pease, Towards a standard upper ontology, in: N. Guarino, B. Smith, C. Welty (Eds.), Proceedings of the International Conference on Formal Ontology in Information Systems. Volume 2001, ACM, New York, 2001: pp. 2–9. doi:10.1145/505168.505170.

[55] R. Mizoguchi, YAMATO: Yet Another More Advanced Top-level Ontology, in: T. Meyer, M. Orgun, K. Taylor (Eds.), Proceedings of the Sixth Australasian Ontology Workshop, 2010: pp. 1–16.

27

[56] S. Borgo, L. Vieu, Artefacts in Formal Ontology, in: A. Meijers (Ed.), Philosophy of Technology and Engineering Sciences, North-Holland, Amsterdam, 2009: pp. 273–307.

[57] V. Mascardi, V. Cordì, P. Rosso, A Comparison of Upper Ontologies. Technical Report DISI-TR-06-21, 2007: pp. 55-64.

[58] S. Semy, M. Pulvermacher, L. Obrst, Toward the use of an upper ontology for US government and US military domains: An evaluation, Technical report MTR 04B0000063, MITRE, 2004.

[59] D. Yuret, The binding roots of symbolic AI: a brief review of the Cyc project, unpublished (1996).

[60] L. Jansen, Categories: The Top-Level Ontology, in: K. Munn, B. Smith (Eds.), Applied Ontology. An Introduction, Ontos, Frankfurt, 2009: pp. 173–196.

[61] A. Rector, Modularisation of domain ontologies implemented in description logics and related formalisms including OWL, in: J. Gennari, B. Porter, G. Yolanda (Eds.), K-CAP 2003 - Proceedings of the 2nd International Conference on Knowledge Capture, ACM, New York, NY, USA, 2003: pp. 121–128.

[62] G.H. Merrill, Ontological realism: Methodology or misdirection?, Applied Ontology. 5:2 (2010) 79–108. doi:10.3233/AO-2010-0076.

[63] I. Johansson, Four Kinds of "Is_A" Relation, in: K. Munn, B. Smith (Eds.), Applied Ontology. An Introduction, Ontos, Frankfurt, 2009: pp. 235–254.

[64] E. Beisswanger, S. Schulz, H. Stenzhorn, U. Hahn, BioTop: An upper domain ontology for the life sciences - A description of its current structure, contents and interfaces to OBO ontologies, Applied Ontology. 3:4 (2008) 205–212. doi:10.3233/AO-2008-005.

[65] S. Schulz, U. Hahn, Towards the ontological foundations of symbolic biological theories, Artif. Intell. Med. 39:3 (2007) 237–250. doi:10.1016/j.artmed.2006.12.001.

[66] D.B. Lenat, Applied ontology issues, Applied Ontology. 1:1 (2005) 9–12.

[67] A. Gangemi, N. Guarino, C. Masolo, A. Oltramari, L. Schneider, Sweetening ontologies with DOLCE, in: A. Gómez-Pérez, V.R. Benjamins (Eds.), Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web, Springer, Berlin, Heidelberg, New York, 2002: pp. 223–233.

[68] B. Smith, W. Ceusters, B. Klagges, J. Köhler, A. Kumar, J. Lomax, et al., Relations in biomedical ontologies, Genome Biology. 6: Suppl 4 (2005) R46. doi:10.1186/gb-2005-6-5-r46.

[69] H. Putnam, It ain't necessarily so, The Journal of Philosophy. 59:22 (1962) 658–671. doi:10.2307/2023596.

[70] S.A. Kripke, Naming and necessity, Blackwell, Oxford, 1980.

[71] V. Gowariker, V.N. Krishnamurthy, S. Gowariker, M. Dhanorkar, K. Paranjape, The Fertilizer Encyclopedia, Wiley, Hoboken, NJ, USA, 2008.

[72] European Commission, Regulation (EC) No. 2003/2003 of the European Parliament of the Council relating to fertilizers, 2003.

[73] J. Paavola, Resources, in: W.A. Darity (Ed.), International Encyclopedia of the Social Sciences. Thomson Gale, 2008. http://www.encyclopedia.com/doc/1G2-3045302263.html (accessed May 13, 2013).

[74] DüMV, Verordnung über das Inverkehrbringen von Düngemitteln, Bodenhilfsstoffen, Kultursubstraten und Pflanzenhilfsmitteln (Düngemittelverordnung - DüMV), 2008.

[75] OBO Download Matrix, (2012). http://www.berkeleybop.org/ontologies/ (accessed October 25, 2012).

[76] Z. Xiang, C. Mungall, A. Ruttenberg, Y. He, Ontobee: A Linked Data Server and Browser for Ontology Terms, in: O. Bodenreider, M.E. Martone, A. Ruttenberg (Eds.), Proceedings of the 2nd International Conference on Biomedical Ontologies (ICBO), Buffalo, NY, USA, 2011: pp. 279–281.

[77] P.L. Whetzel, N.F. Noy, N.H. Shah, P.R. Alexander, C. Nyulas, T. Tudorache, et al., BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications, Nucleic Acids Res. 39:Web Server issue (2011) W541–W545. doi:10.1093/nar/gkr469.

[78] M. d'Aquin, N.F. Noy, Where to publish and find ontologies? A survey of ontology libraries, Web Semantics: Science, Services and Agents on the World Wide Web. 11 (2012) 96–111. doi:doi:10.1016/j.websem.2011.08.005.

[79] E. Paslaru-Bontas, A Contextual Approach to Ontology Reuse: Methodology, Methods and Tools for the Semantic Web, PhD thesis, Free University of Berlin, Germany, 2007.

[80] Z. Xiang, M. Courtot, R.R. Brinkman, A. Ruttenberg, Y. He, OntoFox: web-based support for ontology reuse, BMC Research Notes. 3 (2010) 175. doi:10.1186/1756-0500-3-175.

[81] M. Courtot, F. Gibson, A.L. Lister, J. Malone, D. Schober, R.R. Brinkman, et al., MIREOT: the Minimum Information to Reference an External Ontology Term, Applied Ontology. 6:1 (2011) 23–33. doi:10.3233/AO-2011-0087.

[82] M. Horridge, A Practical Guide To Building OWL Ontologies Using Protégé 4 and CO-ODE Tools, 1.2 ed., The University Of Manchester, Manchester, UK, 2009.

[83] A. Rector, N. Drummond, M. Horridge, J. Rogers, H. Knublauch, R. Stevens, et al., OWL Pizzas: Practical Experience of Teaching OWL-DL: Common Errors & Common Patterns, in: E. Motta, N.R. Shadbolt, A. Stutt, N. Gibbins (Eds.), Engineering Knowledge in the Age of the Semantic Web, Springer, Berlin, Heidelberg, 2004: pp. 63–81.

[84] S. Schulz, D. Seddig-Raufie, N. Grewe, J. Röhl, D. Schober, M. Boeker, et al., GoodOD Guideline v1.0, 2012. http://purl.org/goodod/guideline.

[85] V. Presutti, A. Gangemi, D. Stefano, G.A. de Cea, M.C. Suárez-Figueroa, E. Montiel-Ponsoda, et al., A Library of Ontology Design Patterns, Consiglio Nazionale delle Ricerche (CNR), Roma-Lazio Italy, 2008.

[86] K. Degtyarenko, P. de Matos, M. Ennis, J. Hastings, M. Zbinden, A. McNaught, et al., ChEBI: a database and ontology for chemical entities of biological interest, Nucleic Acids Research. 36:Database issue (2007) D344–D350. doi:10.1093/nar/gkm791.

[87] J. Röhl, L. Jansen, Representing dispositions, Journal of Biomedical Semantics. 2:Suppl 4 (2011). doi:10.1186/2041-1480-2-S4-S4.

[88] HermiT OWL Reasoner, (n.d.). http://hermit-reasoner.com/ (accessed November 14, 2012).

[89] V. Svátek, O. Šváb-Zamazal, V. Presutti, Ontology naming pattern sauce for (human and computer) gourmets, in: E. Blomqvist, K. Sandkuhl, F. Scharffe, V. Svatek (Eds.), Proceedings of the Workshop on Ontology Patterns (WOP 2009), Washington DC, USA, October 25, 2009 (= CEUR 516), http://ceur-ws.org/Vol-516/.

[90] D. Schober, B. Smith, S. Lewis, W. Kuśnierczyk, J. Lomax, C. Mungall, et al., Survey-based naming conventions for use in OBO Foundry ontology development, BMC Bioinformatics. 10:125 (2009). doi:10.1186/1471-2105-10-125.

[91] J. Röhl, L. Jansen, Why functions are not special dispositions: an improved classification of realizables for top-level ontologies, Journal of Biomedical Semantics. 5:27 (2014). doi:10.1186/2041-1480-5-27.

[92] D. Kless, S. Milton, E. Kazmierczak, J. Lindenthal, Thesaurus and ontology structure: Formal and pragmatic differences and similarities, Journal of the Association for Information Science and Technology. 66:7 (2015) 1348–1366. doi:10.1002/asi.23268.

[93] B. Smith, W. Ceusters, Ontological realism: A methodology for coordinated evolution of scientific ontologies, Applied Ontology. 5:3-4 (2010) 139–188. doi:10.3233/AO-2010-0079.

[94] S. Staab, R. Studer, H.P. Schnurr, Y. Sure, Knowledge processes and ontologies, Intelligent Systems, IEEE. 16:1 (2001) 26–34.

[95] M. Uschold, M. King, Towards a methodology for building ontologies, in: Proceedings of the Fourteenth International Joint Conference on AI (IJCAI-95), Montreal, Quebec, Canada, August 20-25, 1995, 1995.

[96] L. Jansen, S. Schulz, The Ten Commandments of Ontological Engineering, in: H. Herre, R. Hoehndorf, F. Loebe (Eds.), Proceedings of the 3rd Workshop of Ontologies in Biomedicine and Life Sciences (OBML), Berlin, 6.-7.10.2011, Leipzig, Germany, 2011.

[97] A. García, K. O'Neill, L.J. Garcia, P. Lord, R. Stevens, O. Corcho, et al., Developing Ontologies within Decentralised Settings, in: H. Chen, Y. Wang, K.-H. Cheung, R. Sharda, S. Voß (Eds.), Semantic E-Science, Springer, U.S.A., 2010: pp. 99–139.

[98] M. Fernández-López, A. Gómez-Pérez, N. Juristo, Methontology: from ontological art towards ontological engineering, in: Proceedings of the Ontological Engineering AAAI-97 Spring Symposium Series, 24-26 March 1997, American Asociation for Artificial Intelligence, 1997.

29

## 9. Appendices

In section 3 we detailed that the steps in our re-engineering method are the results of the practical application of a naïve re-engineering method. The steps in the naïve re-engineering method stem from (a) a theoretical comparison of thesauri and ontologies [42], (b) an analysis of ontology re-engineering literature (Hahn [9], Wroe et al. [11], and Guarino and Welty [17]) and (c) an analysis of general ontology-engineering literature. These steps are summarized in table 4, which also lists the respective authors and publications. We included only steps that are content-focused as well as precise and actionable.

Figure 10 shows the steps of the naïve re-engineering method on the left hand side and relates them to the steps in the final re-engineering method that we introduced in this paper. The relationships indicate that a step in the naïve re-engineering method is either equivalent to the indicated step in the final re-engineering method or that it is a direct or indirect part of the step in the final re-engineering method. It should be noted that figure 10 does not show the various sub-activities of the steps in the final re-engineering method (summarized in Appendix 2).

We did not adopt the OntoClean method [17] from the naïve re-engineering into our final re-engineering method because we did not detect any errors in the is-a hierarchy when applying OntoClean. The reason for not gaining a benefit from OntoClean may be that improving the is-a hierarchy is an implicit result of steps 3, 4 and 5 of our method. In particular, the alignment to a top-level ontology in step 3 may affect the is-a hierarchy in a way that is comparable to an application of the OntoClean method. Nevertheless, the degree of overlap depends on the top-level ontology, but also on a correct application of the top-level ontology and its corresponding set of relationships. It requires further investigation to determine, whether the effects of applying OntoClean are the same as applying our method, or whether OntoClean should be added as an additional step to our method.

Table 4. General steps for the development of qualitatively good ontologies

| General ontology engineering step | Reference backing the step |
|---|---|
| a) Distinction of intensional and extensional entities (universals and particulars in ontological realism) | Smith and Ceusters [93], OBO Foundry principle under discussion, Borgida and Brachman [15] |
| b) Establishment of an is-a hierarchy | Noy and McGuinness [16], Borgida and Brachman [15], Staab et al [94], Hahn [9] [A] |
| c) Alignment to a top-level ontology | Uschold and King [95], Smith and Ceusters [93], OBO Foundry principle under discussion, Jansen and Schulz [96], Hahn [9] |
| d) Application of the Onto-Clean method | Borgida and Brachman [15], Guarino and Welty [17] |
| e) Establishment of a single inheritance hierarchy | Rector [61], Smith and Ceusters [93], OBO Foundry principle under discussion |
| f) Adoption of a well-founded set of ontological relationships that harmonize with the chosen top-level ontology | Borgida and Brachman [15], Accepted OBO Foundry principle |
| g) Definition of a rich set of membership conditions as a basis for the ontology's hierarchy (the is-a relationships) | García et al [97], Noy and McGuinness [16], Borgida and Brachman [15], Staab et al [94], Hahn [9], Wroe et al. [11] |
| h) (Correct) Codification in a formal representation language | Uschold and King [95], Fernández-López et al [98], García et al [97], Borgida and Brachman [15], Staab et al [94], Rector et al [83], Accepted OBO Foundry principle, Hahn [9], Wroe et al. [11] |
| i) Provision of metadata for all classes and relationships such as textual definitions and labels | Accepted OBO Foundry principle, Jansen and Schulz [96] |
| j) Adhering to naming convention for the labels | Accepted OBO Foundry principle, Jansen and Schulz [96] |
| k) Delineation from existing ontologies | Smith and Ceusters [93], Accepted OBO Foundry principle |

[A] Hahn advises to remove cycles in the is-a hierarchy, which is part of the establishment of a (correct) is-a hierarchy.

**Naïve re-engineering method**

a. Identification of thesaurus concepts

b. Identification of thesaurus

c. Identification of whole-part and associative relationships

d. Identification of classes and particulars

e. Establishment of an is-a hierarchy

f. Alignment to a top-level ontology

g. Application of the OntoClean method

h. Establishment of a single inheritance hierarchy

i. Adoption of a well-founded set of ontological relationships that harmonize with the chosen top-level ontology

j. Definition of a rich set of membership conditions

k. Representation in a formal language

l. Provision of metadata for all classes and relationships such as textual definitions and labels

m. Adhering to naming convention for the labels

n. Delineation from existing ontologies

**Final re-engineering method**

1. Preparatory refinement and checking of the thesaurus

2. Naïve conversion

3. Alignment to a top-level ontology

4. Identification of membership conditions

5. Formal specification of membership conditions

6. Adjustment of spelling, punctuation and other aspects of entity labels
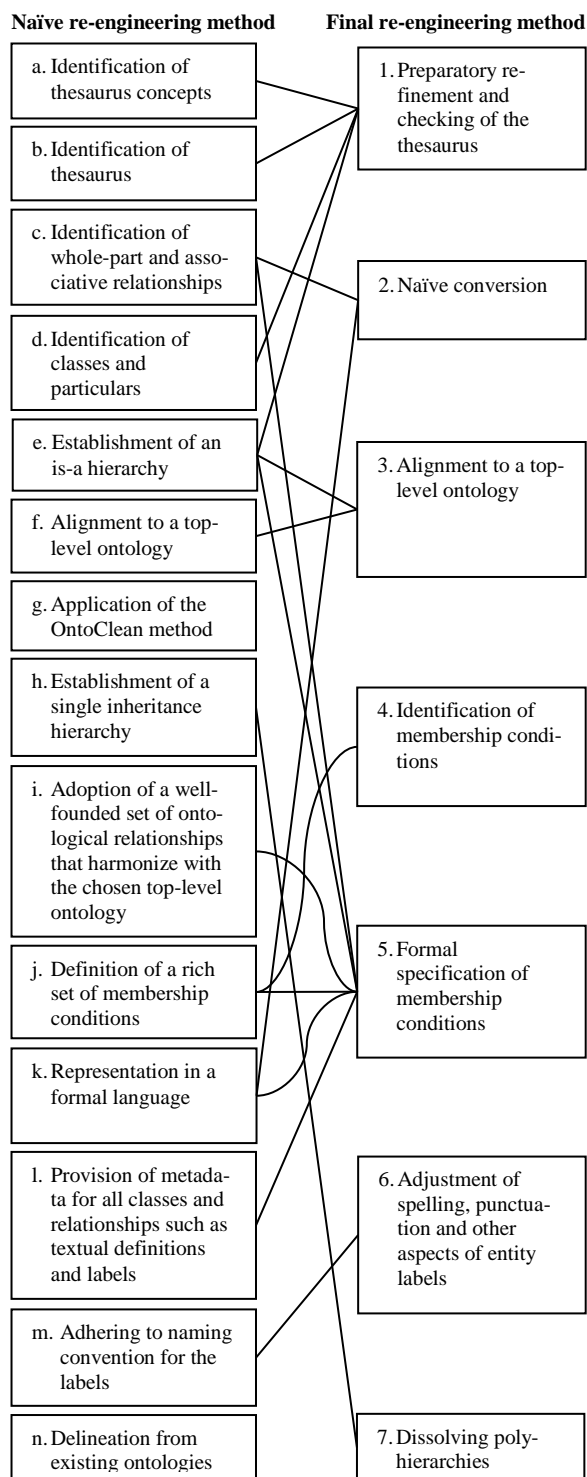
7. Dissolving poly-hierarchies

Figure 10. Relation between steps in the naïve re-engineering method and the final re-engineering method.

*Appendix 2: Overview of the steps and sub-activities of the re-engineering method*

The re-engineering method explained in section 4 consists of various steps and sub-activities. We provide an overview of the sub-activities for every step here.

1. Preparatory refinement and checking of the thesaurus*
   a. Distinction between concepts and terms*
   b. Distinction between different types of hierarchical relationships*
   c. Rejection of invalid relationships*
   d. Removing hierarchical cycles*
   e. Assigning orphans to the thesaurus hierarchy*
   f. Identification of arrays based on characteristics of division*
2. Naïve conversion
   a. Choice or development of conversion tools*
   b. Conversion of the thesaurus into the formal language
3. Alignment to a top-level ontology
   a. Choice of an existing top-level ontology
   b. Alignment of the separated portions of the emerging ontology to the top-level ontology
4. Identification of membership conditions (in natural language)
   a. Identifying the meaning of the classes
   b. Collection of definitions in natural language*
   c. Deciding for principles of including classes and resolving ambiguity*
   d. Extraction or definition of membership conditions
5. Formal specification of membership conditions
   a. Choice of existing domain-specific ontologies and their alignment to the top-level ontology*
   b. Choice of a set of formal relations
   c. "Amendment" of the external ontologies*
   d. Formalization of the class specifications
   e. Adding natural language definitions and comments as class annotations*
   f. Consistency check and inference of class hierarchy
6. Adjustment of spelling, punctuation and other aspects of entity labels*
   a. Choice of a labeling convention*
   b. Adjusting the class labels*
7. Dissolving poly-hierarchies*

* Optional step, the usefulness of which depends on the characteristics of the thesaurus (steps 1, 4.b/c), its storage format and storage system (step 2.a), the availability and quality of existing ontologies (step 5.a/c/e), the intended usage of the ontology (steps 6 and 7), and personal preference in general.

*Appendix 3: Defined arrays of fertilizer concepts in the AGROVOC thesaurus*

This overview presents the arrays that we defined in the course of preparing and checking fertilizer concepts in the AGROVOC thesaurus during the case study. The respective step was discussed in section 4.1. The node labels that indicate the arrays are highlighted in italic font.

*(by type of dominating plant nutrient)*
  Calcium fertilizers
  Magnesium fertilizers
  Nitrogen fertilizers
    Ammonium fertilizers
    Nitrate fertilizers
  Potash fertilizers
  Phosphate fertilizers
    Superphosphate
  Sulphur fertilizers
*(by amount needed by plants)*
  Primary nutrient fertilizers*
    Nitrogen fertilizers
    Phosphate fertilizers
    Potash fertilizers
  Secondary nutrient fertilizers*
    Calcium fertilizers
    Magnesium fertilizers
    Sodium fertilizers*
    Sulphur fertilizers
  Micronutrient fertilizers
    Boron fertilizers*
    Cobalt fertilizers*
    ...
*(by number of plant nutrients)*
  Single nutrient fertilizer
  Compound fertilizers
    Two nutrient fertilizer
      NP-Dünger
      NK-Dünger
      PK-Dünger
    Three nutrient fertilizer
      NPK-Dünger
*(by nutrient release time)*
  Fast release fertilizers*
  Slow release fertilizers
*(by substance group)*
  Organic fertilizers
    Biofertilizers
    Compost
    Fish manure
    Green manure
    Guano
  Inorganic fertilizers
  Organomineral fertilizers
*(by aggregate state)*
  Solid fertilizers*
  Liquid fertilizers
  Liquid gas fertilizers

* Concept added, i.e. not included in AGROVOC

*Appendix 4: Membership conditions after alignment (step 4)*

This appendix summarizes the membership conditions for fertilizer concepts and fertilizer-related concepts in the AGROVOC thesaurus. The membership conditions are fundamentally based on their extraction from natural language definitions as described in section 4.3. Nevertheless, the status presented here was only achieved after the alignments of the fertilizer classes and adopted ontologies to a top-level ontology. This was elaborated in section 4.3.

In addition, table 6 indicates the plant nutrient levels extracted from the fertilizer regulation by the European Commission [72] and the German fertilizer regulation "Düngemittelverordnung" [74].

In tables 5–7, we use natural language to describe the membership conditions. The wording is as close as possible to the class names, relationship names and property names of the imported ontologies. We have split the complex conditions of some classes ('Fish manures' and 'Guano') into several dependent conditions using some auxiliary classes (indicated in italic font) in order to improve the readability. These classes do not appear in the formal specification where they are simply nested into each other, that is, the name of the auxiliary classes is replaced by their respective definitions.

Table 5. Membership conditions of fertilizer classes

| Class/fertilizer type | Membership conditions | Necessary/sufficient[A] |
|---|---|---|
| Fertilizers | being a compound of collective material entities<br>bearing the disposition to release plant nutrients<br>having a component that has a minimal mass proportion of 1680 ppm plant nutrients as granular part | necessary |
| fertilizer types listed in table 6, e.g., calcium fertilizer | being a fertilizer<br>having a component that has the minimal mass proportion of a plant nutrient (chemical atom) as granular part as indicated in table 6, e.g., the mass proportion of 143,000 ppm calcium bound in some molecule containing calcium | see table 6 |
| Compound fertilizers | being a fertilizer<br>having a component that has minimal mass proportion of 2729 ppm of two or more different primary plant nutrients (nitrogen, sulphur or potassium) as granular part. | necessary |
| Micronutrient fertilizers | being a fertilizer<br>having a component that has a mass proportion of 1680 ppm plant micronutrients as granular part | necessary |
| Organic fertilizers | being a fertilizer<br>having a component that has a significant mass proportion of a carbon-based molecule as granular part | necessary |

32

| Class/fertilizer type | Membership conditions | Necessary/sufficient[A] |
|---|---|---|
| Biofertilizers | being a fertilizer<br>being the outcome of a fixing/binding process or a solubilizing process in which the agent is some living organism and the patient has plant nutrients as granular part | necessary and sufficient |
| Composts | being a fertilizer<br>being the outcome of a decomposition process in which the agent is some living organism and the patient is a dead body | necessary and sufficient |
| Fish manures | being a fertilizer<br>being the outcome of a crushing or powdering process in which the patient is 'dried fish rest';<br>'dried fish rest' is defined as the outcome of a drying process in which the patient is 'fish rest';<br>'fish rest' is being defined as the dead body of fish or physical parts thereof | necessary and sufficient |
| Green manures | being a fertilizer<br>being the outcome of a decomposition process in which the agent is a living organism and the patient is the dead body of a plant or a physical part thereof | necessary |
| Guano | being a fertilizer<br>being the outcome of a decomposition process in which the agent is a living organism and the patient is 'specific excrements'; 'specific excrement' refers here to the outcome of the excretion action in which the agent is a seabird or fish or goat or bat or whale | necessary |
| Inorganic fertilizers | being a fertilizer<br>not being an organic fertilizer | necessary |
| Organo-mineral fertilizers | being a fertilizer<br>having some organic fertilizer as a component<br>having some inorganic fertilizer as a component | necessary |
| Liquid fertilizers | being a fertilizer<br>being a liquid material | necessary and sufficient |
| Liquid gas fertilizers | being a fertilizer<br>being a gaseous material | necessary and sufficient |
| Slow release fertilizers | being a fertilizer<br>bearing the disposition to release plant nutrients slowly | necessary and sufficient |
| fertilizer pesticide combinations | being a compound of collective material entities[B]<br>having a significant mass proportion of fertilizer as component<br>having a significant mass proportion of pesticide as component | necessary |

[A] Classification as primitive class specified with necessary conditions or as defined class specified with necessary and sufficient conditions
[B] This condition has been amended in line with the 'fertilizer' specification.

Table 6. Necessary parts of element- or molecule-focused fertilizers in relation to ChEBI

| Class/fertilizer type | Granular part (as defined in ChEBI) | pm of atom | Necessary/sufficient[A] |
|---|---|---|---|
| Calcium fertilizers | 'calcium molecular entity' | 143000 Ca | necessary and sufficient |
| NPK fertilizers | 'phosphorus molecular entity' and 'potassium molecular entity' and 'nitrogen molecular entity' | 654 P<br>2075 K<br>5000 N | necessary and sufficient |
| Nitrogen phosphorus fertilizers | 'phosphorus molecular entity' and 'nitrogen molecular entity' | 654 K<br>5000 N | necessary and sufficient |
| Nitrophosphates | 'calcium hydrogen phosphate' and 'ammonium nitrate' and 'diammonium hydrogen phosphate' | n/a | necessary |
| Nitrogen potassium fertilizers | 'potassium molecular entity' and 'nitrogen molecular entity' | 2075 K<br>5000 N | necessary and sufficient |
| Phosphorus potassium fertilizers | 'phosphorus molecular entity' and 'potassium molecular entity' | 654 P<br>2075 K | necessary and sufficient |
| Magnesium fertilizers | 'magnesium molecular entity' | 84600 Mg | necessary and sufficient |
| Phosphate fertilizers | 'phosphorus molecular entity' | 30956 P | necessary and sufficient |
| Rock phosphate | 'apatite' [B] | n/a | necessary |
| Superphosphate | 'calcium sulfate' and 'calcium bis(dihydrogenphosphate)' | n/a | necessary |
| Potash fertilizers | 'potassium molecular entity' | 58100 K | necessary and sufficient |
| Sulphur fertilizers | 'sulfur molecular entity' | 55000 S | necessary and sufficient |
| Nitrogen fertilizers | 'nitrogen molecular entity' | 45000 N | necessary and sufficient |
| Ammonium fertilizers | 'ammonium compound' | n/a | necessary |
| Nitrate fertilizers | 'nitrates' | n/a | necessary |

[A] Classification as primitive class specified with necessary conditions or as defined class specified with necessary and sufficient conditions
[B] 'Apatite' is subsumed under 'collective material entity' in ChEBI, so that no reference to the granular part is necessary.

Table 7. Membership conditions of classes closely related to agricultural fertilizers

| Class | Membership conditions | Necessary/sufficient[A] |
|---|---|---|
| plant nutrient | (a) being a molecular entity (b) being either a primary plant nutrient or secondary plant nutrient or plant micronutrient (c) bearing the disposition to be picked up by plants | necessary and sufficient |
| plant micro-nutrient | (a) being a plant nutrient (b) being a molecule that contains either boron or copper or iron or manganese or molybdenum or zinc | necessary |
| primary plant nutrient | (a) being a plant nutrient (b) being a molecule that contains either phosphorus or potassium or nitrogen | necessary and sufficient |
| secondary plant nutrient | (a) being a plant nutrient (b) being a molecule that contains either calcium or magnesium or sulfur | necessary |
| plant nutrient disposition | (a) being a disposition (b) being realizable by a plant nutrient uptake process | necessary |
| plant nutrient uptake process | (a) being a kind of bio molecular process, the locus of which is a plant and the participants in the process are plant nutrients (b) realizing some disposition of being a plant nutrient | necessary |
| plant nutrient release disposition | (a) being a disposition (b) being realizable by a plant nutrient release process | necessary |
| plant nutrient release process | (a) being a process (b) realizing some disposition to release plant nutrients | necessary |
| plant nutrient slow release disposition | (a) being a plant nutrient release disposition | necessary |

[A] Classification as primitive class specified with necessary conditions or as defined class specified with necessary and sufficient conditions

The table shown here presents the result of adjustments to the class labels in the course of the re-engineering of the fertilizer concepts and fertilizer-related concepts in the AGROVOC thesaurus. The step was described in section 4.6.

Table 8. Comparison of class labels with former preferred terms in the AGROVOC thesaurus

| Preferred term in the AGROVOC thesaurus | Class label in the fertilizer ontology |
|---|---|
| Fertilizers | portion of fertilizer |
| Nitrogen fertilizers | portion of nitrogen fertilizer |
| ammonium fertilizers | portion of ammonium fertilizer* |
| nitrate fertilizers | portion of nitrate fertilizer* |
| Phosphate fertilizers | portion of phosphate fertilizer |
| Rock phosphate | portion of rock phosphate fertilizer* |
| Superphosphate | portion of superphosphate fertilizer * |
| Potash fertilizers | portion of potash fertilizer |
| Calcium fertilizers | portion of calcium fertilizer |
| Magnesium fertilizers | portion of magnesium fertilizer |
| Sulphur fertilizers | portion of sulphur fertilizer |
| Compound fertilizers | portion of compound fertilizer |
| NPK fertilizers | portion of NPK fertilizer |
| Nitrogen phosphorus fertilizers | portion of nitrogen phosphorus fertilizer |
| Nitrophosphates | portion of nitrophosphate fertilizer * |
| Nitrogen potassium fertilizers | portion of nitrogen potassium fertilizer |
| Phosphorus potassium fertilizers | portion of phosphorus potassium fertilizer |
| Micronutrient fertilizers | portion of micronutrient fertilizer |
| Organic fertilizers | portion of organic fertilizer |
| Biofertilizers | portion of biofertilizer |
| Composts | portion of compost fertilizer |
| Fish manure | portion of fish fertilizer |
| Green manures | portion of green manure fertilizer |
| Guano | portion of guano fertilizer |
| Organomineral fertilizers | portion of organomineral fertilizer |
| fertilizer pesticide combinations | portion of fertilizer pesticide combination |
| Inorganic fertilizers | portion of inorganic fertilizer |
| Liquid fertilizers | portion of liquid fertilizer |
| Liquid gas fertilizers | portion of liquid gas fertilizer |
| Slow release fertilizers | portion of slow release fertilizer |
| seabirds | seabird |
| Goats | goat |
| whales | whale |
| plant | plant |
| degradation | degradation |
| solubilization | solubilization |
| crushing | crushing |
| drying | drying |
| Excretion | excretion |
| pesticides | pesticide |