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

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 syntactically only and ignore the problems that stems from interpreting vocabularies as statements of truth (ontologies). Current reengineering methods also do not make use of the semantic capabilities of formal languages like OWL in order to detect logical mistakes and to 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. The method results in a semantically adequate ontology that does not only contain a semantically rich description of the entities to be modeled, but also enables non-trivial consistency checks and classifications based on automated reasoning, and can be integrated with other ontologies following the same development principles. The identification of membership conditions, the alignment to a top-level ontology and formal relations, and the consistency check and inference using a reasoner are the central steps in our method. We explain the motivation and sub-activities for each of these steps and illustrate their application through a case study in the domain of agricultural fertilizers based on the ACROVOC Thesaurus. Foremost, our method shows that simple syntactic conversions are insufficient to derive an ontology from a thesaurus. Instead, considerable structural changes are required to derive an ontology that corresponds to the reality it represents. Our method relies on a manual development effort and is particularly useful where a highly reliable is-a hierarchy is crucial.

Keywords*:* Thesaurus Re-engineering, Ontology development

1. Introduction

In information science, ontologies are statements of necessary truth about the common features of entities in reality in a computable formal language. The use of a formal system supports automated reasoning, which comprises not only an automated consistency check of the ontology (i.e. proving the absence of contradictions), but also the inference of new facts that have not explicitly been asserted [1].

The creation of knowledge-dense ontologies can take tremendous time [2]. For this reason it is desirable to re-use existing models as ontologies [3]. 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 incorrectly referred to as terminologies, are examples of non-ontological resources and are generally considered interesting candidates for re-use as ontologies [4], [5]. The reason is that such vocabularies have often matured over decades and contain several thousand up to hundreds of thousands of concepts and natural language terms. This eliminates or at least reduces the effort of eliciting concepts in the ontology development process. Second, the concepts in a vocabulary are generally structured through a number of relationships. These relationships can be used as a starting point for developing the structure of an ontology.

There are divergent opinions of what is necessary for the re-use of a vocabulary as an ontology. Some methods suggest that the re-use requires mainly a syntactical change by describing the data model as

well as the content of a thesaurus in a logic-based language [4], [5]. Other approaches point out that ontologies make finer distinctions between relationships than vocabularies [6]. Still others point at the need for fundamental structural changes in order to derive an ontology from a vocabulary [7]–[9]. Finally, there are authors who emphasize the need for applying philosophical principles to build ontologies, particularly emphasizing the stance of ontological realism [10], [11].

The divergence of these opinions stems from different views on what ontologies and formal languages are. Many methods for reengineering vocabularies into ontologies [4], [5], [12]–[15] describe the "ontology" in the Resource Description Framework (RDF) [16], which is standardized by the World Wide Web Consortium (W3C) and often called a "Semantic Web Standard". Unlike the authors that use RDF for reengineering vocabularies into ontologies we do not consider RDF to be a language that is adequate for representing ontologies in the first instance. The main reason is that RDF, specifically the RDF Schema [17], does not strictly separate between classes and their instances and subsequently does not facilitate reasoning and the logically correct integration of independently developed ontologies—a prerequisite that is essential to achieve visions of a Semantic Web as it was expressed by Berners Lee et al. [18]. Not separating classes and their instances must also be considered the main reason, why RDF with its formal semantics for RDF [19] is computationally intractable [20] and unlikely to ever have complete reasoning support [21, Sec. 1.3].

What we consider actual formal languages for the representation of ontologies are languages that are based on first order logic, description logic or modal logic. The Web Ontology Language (OWL) [22] with its description logic semantics [23], [24] is an example of a formal language that strictly separates instances (individuals) and abstractions of them (classes). OWL is another Semantic Web standard and the recommended ontology language of the W3C. The computational tractability, the strong reasoning support, as well as the use of XML-like syntaxes and unique identifiers (IRIs/URIs) are considerable advantages and a reason for the high popularity of OWL.

Because there are mappings from OWL to RDF and vice versa [25] as well as an RDF-based Semantics [26] for OWL, the distinction between OWL and RDF appears to have become blurred for many people and the use of RDF is considered as a "Semantic Web representation" or "RDF/ OWL representation" of ontologies [4], [5], [12]–[15]. This may be described as the widespread understanding of ontologies in the (not clearly defined) Semantic Web community.

The blurring of RDF and OWL is fatal from the perspective of those, who model ontologies using OWL and who respect the description logic semantics. While it is true that ontologies described in OWL—just as literally any data or datamodel—can be syntactically translated into RDF descriptions, it is a wrong assumption that RDF descriptions can always be interpreted as OWL descriptions of ontologies. Such interpretation requires that what the structure of what is described in RDF complies with the description logic semantics of OWL. This is not the case for many of the so-called ontologies described in RDF that result from applying current reengineering methods to vocabularies [4], [5], [12]–[15]. As we will show in this paper, reengineering vocabularies into ontologies using OWL changes the structure of vocabularies considerably. We believe that only based on these structural changes visions such as the one of the Semantic Web can become true. It is only ontologies using OWL that give hope for integrating independently developed ontologies in a logically consistent way and with correspondence to the represented reality.

We are not aware of any method that explicitly describes the reengineering of vocabularies into ontologies using OWL, although such methods are implicitly applied by at least some groups that develop ontologies in the OBO Foundry [27]. We will further discuss existing reengineering at the end of this paper (in section 5), because the uniqueness of our method and contribution and how it differs from existing methods is more understandable once our method is fully laid out. The current lack of explicit methods that guide the (re-)engineering of proper ontologies is a major obstacle for achieving visions like the Semantic Web or integrating at least ontologies in the same subject area.

The goal that we pursue in this paper is to lay out a method for the reengineering of vocabularies into ontologies using the formal language and Semantic Web standard OWL. The re-engineering method that we present is instructive and *content focused* so that it can be easily applied. We take the content of both thesauri and ontologies to comprise (a) their structure, (b) their syntactic specification and (c) the labeling of their structural elements. The structure includes (b_1) the representational units (otherwise called "concepts", "classes", "terms" or "entities") and (b_2) the relationships between these units (also called "formal

relations" or "object properties"). Further, our method aims at developing a *semantically adequate* ontology that

a) makes full use of the semantic expressivity of OWL,
b) can be integrated with other ontologies following the same development principles, and
c) is consistent and provides reasoning results that correspond to the represented reality.

The method that we are going to present guides specifically the re-engineering of a *thesaurus*, a specific type of vocabulary. The reason why we focus on the reengineering of thesauri is that there are structural differences between different types of vocabularies (e.g. simple lists of terms, thesauri, taxonomies or classification schemes [28]) and their reengineering may differ. The thesaurus is a well-defined type of controlled and structured vocabulary [29], [30] and there exist presumably several hundreds of thesauri that could be adopted as ontologies [31]. Our method has thus potential to be applied to many existing vocabularies. We will demonstrate the validity of our method by applying it to re-engineer a portion of a specific thesaurus, namely the fertilizer branch of the AGROVOC Thesaurus [32].

The paper is structured as follows: In subsequent section 2 we will detail how the re-engineering method was derived. Section 3 will introduce the steps of our re-engineering method. In an earlier paper [33] we provided an outline of this method and present here a matured version in more detail. In section 4 we will reflect on the method as a whole. It is only in the end of this paper, in section 5, when we will explain, how our method differs from existing reengineering methods. The reason for this sequence is that understanding our method will help understanding its difference from existing reengineering methods that are based on RDF-oriented and other understandings of ontologies. Section 6 concludes the paper.

2. 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 naive 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 [30] with the structure of realist ontologies [34] and their specific representation in the description logic OWL [35, p. 2], [36]. 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. We did, however, not find any single method comprising all the steps that we have adopted. This inclusion of steps from ontology engineering partially explains why re-engineering a thesaurus into an ontology is more than a syntactic conversion of a thesaurus: These steps are not part of thesaurus development and sometimes not even possible to implement in thesauri that adhere to ISO 25964-1:2011. The combination of the steps from the theoretical analysis and the general ontology engineering literature constituted the naive re-engineering method and is laid out in Appendix 1.

In the second phase of refining and validation, we applied the naive 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 naive 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, we expected the semantically adequate re-engineering of a thesaurus into an ontology to be highly time-consuming, which turned out to be true. This 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 are 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 about 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: acriculture). 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 [32] which comprises 31 concepts subordinated to 'Fertilizers'. In addition, 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 of a project participant in a fertilizer ontology, but also because the AGROVOC is a mature and widely used thesaurus.

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

Our re-engineering method consists of seven steps that are shown in figure 1. The arrows connecting the steps indicate that the method is expected to be applied iteratively. Appendix 2 provides a more detailed overview of the method by summarizing the subactivities for each step. The following subsections will, for each of the steps, discuss the purpose, provide an explanation of the activities involved and finally 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 subactivities that we will introduce in the explanation of the step.

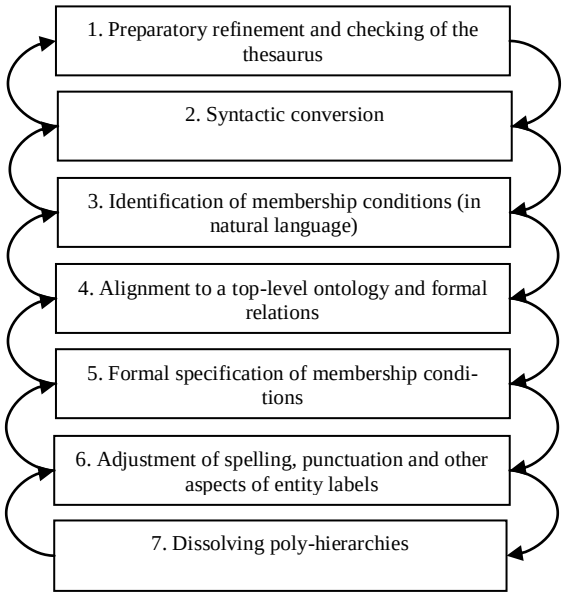

Figure 1. Method for engineering quality ontologies based on thesauri

3.1. Step 1: Preparatory refinement and checking of the thesaurus

Purpose

We base our re-engineering method on the thesaurus standard ISO 25964-1:2011 [30]. Thesauri in practice 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. Thus, domain-specific thesauri may often not have adopted the past or recent 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. Further, applying optional features of a thesaurus like the node labels for indicating characteristics of division of the thesaurus concepts are helpful for later analytical steps; for this reason we encourage them here.

In some cases, the refinement of the thesaurus may be impeded by the specific thesaurus management software in place. For this reason, this methodical 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 to derive a useful basis for the is-a hierarchy of an ontology.

Actions to be taken

The following things should be ensured in a thesaurus in accordance with the ISO thesaurus standard ISO 25964-1:2011:

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" [30, Sec. 2.11], and **terms**, "words or phrases used to label a concept" [30, Sec. 2.61], is explicit in the data model in the thesaurus standard ISO 25964-1:2011. If a thesaurus does not make this distinction, then concepts needs to be created that represent the preferred terms and their respective bundle of non-preferred terms. Eventual corrections should generally be automatable. Attention should be paid as to whether there exist hierarchical or associa-

tive relationships, which relate one or two non-preferred terms. Such relationships would be considered erroneous in term-based thesauri and should be "transferred" to concept-to-concept relationships, just like the relationships between preferred terms. 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 different relationships 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. 'bicylce wheel' and 'bicycle') and (3) the **instance relationships** between a general concept and an instance (e.g. 'Mountains' and 'Alps') [30, Sec. 10.2.2]. 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 the hierarchical relationships there may also be detected relationships that are not conformant with the semantics of the relationship defined in the thesaurus standards and should not be transferred into the ontology. There may be paid less attention to the correctness of **associative relationships**. These relationships are used for "suggesting additional or alternative concepts for use in indexing or retrieval" [30, Sec. 10.3]. They are to be applied between "semantically or conceptually" related concepts that are not hierarchically related [30, Sec. 10.3]. Associative relationships can be ignored at this stage, because their usefulness in ontologies will be critically assessed in step 4.

(d) The thesaurus should also be analyzed for **cyclic hierarchical relationships**. Such cycles are considered erroneous in thesauri and cannot be accepted in the ontology as well, since they bear a logical contradiction. Cycles are best addressed in connection with step 4 of our method.

(e) **Orphans**, 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 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 guidance for "facet analysis" for the identification of node labels [37], [38, p. 5.2], the activity remains an intellectual one for which no proper guidance is available.

Thesauri may contain further kinds of errors such as one-directional 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 [30, Sec. 14.3]. We will not further discuss such errors here.

Application of the step to the fertilizer ontology

(a) The AGROVOC does not distinguish between concepts and terms. Unique identifiers (term codes) are provided for terms only, not concepts. A transformation as shown in figure 2 was done 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 syntactic conversion (step 2).

(b) As with many thesauri, AGROVOC does not distinguish between different types of hierarchical relationships. But, as it happens, our analysis revealed that all hierarchical relationships between 'fertilizer' and its subordinated concepts are proper generic relations. Other parts of the AGROVOC thesaurus do in fact display the other types of hierarchical relationships in thesauri like the instance relationship

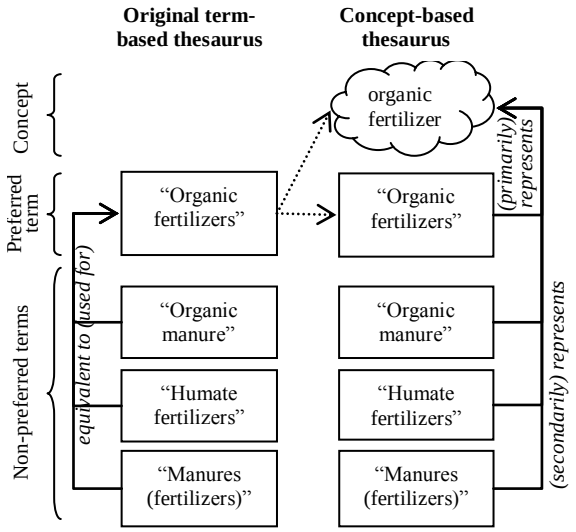

Figure 2. Conversion process from a term-based thesaurus like the AGROVOC to a concept-based thesaurus

(Colorado River—Rivers) or the hierarchical part-of relationship (Root hairs—Roots).

We noted some erroneous relationships amongst the fertilizer-related concepts. 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 they will not be transferred into the ontology as we motivated in subsection 3.2. We did not encounter relationships using a non-preferred term as a relatum that we would have to consider as structural relationship in the ontology, and we only found one situation where a scope note was provided for a non-preferred term. In this case we simply assigned the scope note to the concept, because there was no scope note for the preferred term ('organic fertilizer').

(c–f) We could not detect any *hierarchical cycles* in the hierarchy. Also the detection of *orphans* did not play any role in our case study. The AGROVOC thesaurus does not contain any *node labels indicating characteristics of division*. We were, however, able to define several of them grouping kinds of fertilizers such as the type of dominating plant nutrient, the number of plant nutrients, or the release time of plant nutrients. The complete list of defined arrays with their respective node labels is provided in appendix 3.

Our analysis revealed that the 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.

3.2. Step 2: Syntactic conversion

Purpose

Syntactic conversion aims at representing the thesaurus in a formal language so that it can be further modified in an ontology editor. Further, the formal representation allows the unambiguous interpretation of the ontology, the use of automated reasoning tools to check the ontology for consistency (the absence of contradictions from the joint assertions made in an ontology [39, p. 538]) and to infer the class hierarchy in later steps, but also to exchange the ontology in a common format. It is well possible that the model resulting from the syntactic conversion shows inconsistencies and contradictions that can be detected using automated reasoning. The correction of these inconsistencies and contradictions is the subject of forthcoming methodical steps.

Actions to be taken

Three actions may be distinguished in this step:

a. Choice of a formal language
b. Choice or development of conversion tools
c. Conversion of the thesaurus into the formal language

(a) While, in principle, a choice between formal languages can be made, we focus on the popular OWL in its 2nd version [22] in combination with its "direct semantics" [23] that builds on description logic. An advantage of OWL is that there exist various reasoning algorithms for consistency checking and generating the inferred class hierarchy (explained in more detail in step 5).

(b) 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 possibility to use existing tools instead of developing custom scripts or programs is higher, if the thesaurus is available in common exchange formats such as SKOS [40].

(c) After the refinement of the thesaurus 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 3. The diagram is to be read as follows: some concepts (in thesauri) reference indi-

viduals (in OWL). The name of the relation (*in italic*) expresses the meaning of the relation in the indicated direction.

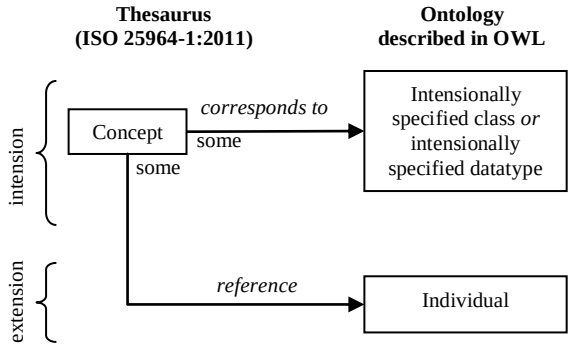

Figure 3. Relatedness of the relata in thesauri and the relata in OWL

A thesaurus concept, as well as facets in their use as top-level elements, can either correspond to an intensionally specified class or an intensionally specified datatype. The terms of a thesaurus and the labels of the facets become labels of classes. Thesaurus concepts can also reference extensional entities such as individuals (e.g the Yangtze River) or specific collections of individuals (e.g. the Rocky Mountains as a specific collection of mountains). Language tags allow distinguishing 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 custom subtypes of such. It might also be desirable to transform node labels into "housekeeping classes" that serve for ontology 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 can be considered as a workaround that is needed because OWL does not provide a modeling primitive corresponding to node labels that can be used for this purpose.

Figure 4 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 or (rather uncommonly) a data subproperty axiom in OWL. Nevertheless, the is-a relationships are preliminary and can become subject of smaller or more fundamental changes in connection with steps 3 and 4.

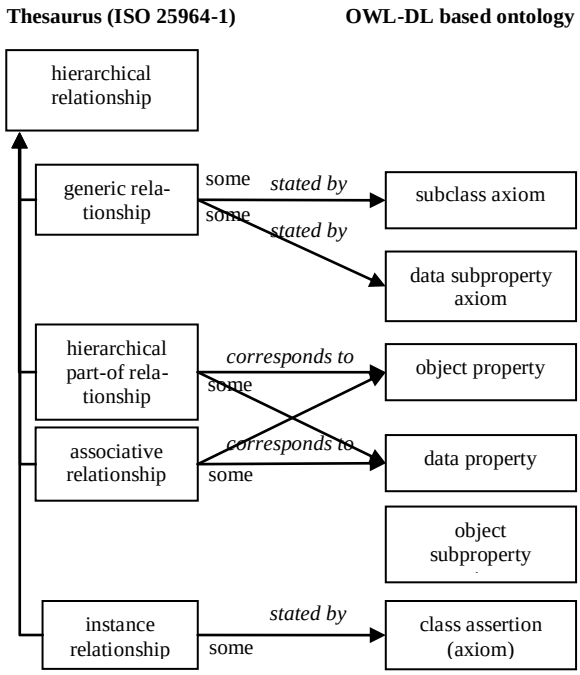

Figure 4. Relatedness of relationships in thesauri and relationships in OWL

Hierarchical whole-part relationships in thesauri should be tentatively modeled as unspecific part-of relationships and represented by object properties or (less commonly) data properties in OWL. The relationships are subject to potential further refinement depending on the set of formally defined relationships that shall be adopted (see step 4). Moreover, the hierarchical whole-part relationships as well as other relationships are subject of validity assessment in step 3 (they must be membership conditions of the classes that they connect).

The **instance relationships** in thesauri may correspond to relationships between an individual and a class—an assertion that is generally not considered part of the ontology, but rather of a knowledge base. As such it is to be rejected as part of an ontology, acknowledging that knowledge bases can be represented by OWL as well. Instance relationships in a thesaurus are then expressed by class assertion axioms in OWL.

Associative relationships *may* give hints that there is an ontological relationship between two concepts that contributes to one concept's formal specification as an ontology class. We recommend checking the usefulness of associative relationships after step 3 rather than converting them directly into relationships in the ontology here. The associative relationships, just like the hierarchical whole-part relationships, must be membership conditions of the classes that they connect in order to be validly applied in the ontology. In our case study they turned out to be invalid ontology relationships in all cases. The associative relationships also need to be refined in order to correspond to any relationship in ontologies and are then represented by object properties or (less commonly) data properties in OWL. Since modeling relationships between relationships is not subject of thesaurus work, there will be no use of the object subproperty axioms to assert generic relationships.

Application of the step to the fertilizer ontology

It turned out to be not useful to follow the actions described for this step in the case of the AGROVOC thesaurus. The reason is that the effort for an automated syntactic conversion would have been much bigger than the manual conversion that we pursued in the end. 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 is far too large to be processed efficiently. It required a computer with 8 processing cores and 8 GB of free memory to even load the file in a reasonable amount of time. 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 just very few classes are subordinated by the subclass axiom. We wanted to start with the class hierarchy as it is presented in the original AGROVOC thesaurus, though. The even bigger problem is that the class labels were not attached to the classes in a way that Protégé could display the class labels.

For these reasons, and since we wanted to reengineer a relatively small portion of the AGROVOC thesaurus only, it was faster for us to enter the class hierarchy for the 'fertilizers' tree manually using the Protégé-OWL editor that we will also use for the formal specification of classes in step 5. We started the conversion with 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 into preferred and non-preferred terms by assigning them to the annotation properties "preferred term" and "non-preferred term" respectively, which we newly defined as subproperties of the default property "label". We also copied the preferred term to the "label" annotation property where it will later be subject to further modification (see step 6). Further, 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. The terms and notes in languages other than English were omitted when entering the thesaurus terms as class labels. Finally, we organized the class hierarchy (the is-a hierarchy) in the ontology in precisely the same way as they could be found in the AGROVOC thesaurus.

3.3. Step 3: 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 *membership conditions*. The goal is to identify as complete as possible characteristics that can act as *necessary* membership conditions, because they are valuable for checking the consistency of the is-a hierarchy and to infer class subsumptions automatically. It is also desirable to identify *necessary and (jointly) sufficient* membership conditions that *define* a class, because it is only defined classes under which other classes can be subsumed by automated reasoning. Nevertheless, one also needs to be aware that wrongly stated membership conditions may result in the mistaken exclusion of real-life entities and/or wrong reasoning results. 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 prepares the ground for later alignment (step 4) and formal specification of the classes and their membership conditions (step 5).

Actions to be taken

Two actions may be necessary in this step:

a. Collection of definitions in natural language

b. Extraction or definition of membership conditions

The most fundamental step in the definition of membership conditions is to have a clear idea of which types of things are to be modeled as classes in the ontology to be developed. For this purpose, we exploit all the means that (at least in principle) a thesaurus offers to express the meaning of its concepts (assigned natural language terms, hierarchy, associative relations, qualifiers, scope notes, definitions). As ISO 25964-1:2011 neither considers definitions necessary nor offers any rules for definitions, many thesauri do not contain any. For this reason it may often be desirable to collect natural language definitions from other sources to become aware of possible ambiguities of concept meanings, but also because they may contain criteria that can be adopted as membership conditions. These encyclopaedias and dictionaries should be as subject-specific as possible in order to have a qualitatively good basis for the definition of membership conditions. Where there are no useful encyclopaedia or dictionary definitions it may be necessary to consult domain experts to create explicit definitions. Any definition needs to be in line with the meaning of a thesaurus concept.

Specifying membership conditions may appear trivial at first sight, but it isn't. It may, in fact, lead to comprehensive investigations and face the ontology developer with difficult decisions. For example, one will generally have an intuitive idea of what a concept labeled "water" represents. If being asked, 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. Terms in natural language are almost always ambiguous and have different meanings in different communities and cultural contexts. Sometimes the terms have even multiple meanings in a single community, particularly if there are different schools of thought. In such cases, an ontology may need to contain several classes for a given term, each for every meaning.

There exists little practical guidance for deciding whether or not (a) a membership condition is a valid (necessary) membership condition and (b) one or more membership conditions constitute a set of *jointly sufficient* membership conditions for a given kind of entity. 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 [41], [42, pp. 119–122], [43, pp. 35–36]. 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 is then not considered a calcium fertilizer, even if it misses the minimum amount just slightly. At this point it is also useful to check, if the hierarchical whole-part relationships or the associative relationships in the thesaurus can be adopted as valid membership conditions.

There may also be kinds of entities for which it is simply not possible to define any membership condition. In such cases, natural language definitions should be provided, which do not need to refer to membership conditions, but may provide examples or *typical* characteristics. Natural language definitions are in any case helpful for both ontology maintainer and user. Examples or explanations of common misunderstandings of what a kind of entity encompasses should be included in comments, not in definitions.

Application of the step to the fertilizer ontology

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

We compensated the lack of definitions in the AGROVOC thesaurus by encyclopedic and regulatory definitions. More specifically, we obtained the definitions primarily from *The Fertilizer Encyclopedia* [44] and a fertilizer-related regulation by the European Commission [45]. While they covered most fertilizer classes, we sometimes had to use definitions from other sources or had to create custom definitions using the advice of subject experts. 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 by the European Commission about the concept 'fertilizer'. These information form the basis for our analysis. The hierarchical context of 'fertilizer' in the AGROVOC thesaurus and a dictionary definition of 'resource' [46] suggest that fertilizer is understood as an input to farming in the AGROVOC thesaurus, farming being a kind of value production. Nevertheless, the fertilizer-hierarchy does not support the assumption that fertilizers are truly included as products, e.g., by considering the fertilizer packaging. Our assumption is rather that fertilizers are referred to with respect to their scientific functioning in the agricultural domain—without taking account of its social contexts—and we follow this understanding, which corresponds to the definitions in *The Fertilizer Encyclopedia* and the fertilizer-related regulation by the European Commission.

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 [44]	*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 [45]	*Fertiliser:* material, the main function of which is to provide nutrients for plants.

The encyclopedia definition as well as the definition by the EC commission 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 (that are roots), so that the nutrients do not have to go the chemical reaction path via the soil. For this reason we re-formulated 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 case of fertilizers applied on soils and a property of 'being liquid' in case of fertilizers applied on plant leaves. This requires detailed further investigation, which we did not pursue.

The formulation of condition (c) is not satisfactory as well. 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 analysed the fertilizer-related regulation by the European Commission [45] and the official regulation in Germany, the "Düngemittelverordnung" [47], 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 that contain the lowest proportions of plant nutrients (plant micronutrients): a minimal mass proportion of 0.17 %. It is the minimum requirement that we can adopt for fertilizers as necessary condition (c):

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

This condition cannot contribute to a specification of fertilizers with necessary and sufficient conditions,

because the condition (c) in combination with the other conditions is also true for a lot of water-soluble substances with little amounts of any plant nutrient (e.g. nitrogen) that would not be considered fertilizers, e.g. various medicaments. Fertilizers can thus be characterized with necessary conditions only. This circumstance made us wonder, whether it is invalid to interpret "significant amounts" of plant nutrients with an *absolute* minimum amount of plant nutrients. One may be more successful to identify a *relative* minimum amount of plant nutrients for fertilizers. This requires further investigation that we did not pursue here.

Specific fertilizer types

In the way we analysed 'fertilizer' in general, we also analysed the other fertilizer types for their meaning and their membership conditions. All of them have one fundamental membership condition—being a fertilizer—and thus inherit all membership conditions from 'fertilizer'.

We faced similar problems like with the class 'fertilizer' when identifying membership conditions for the classes 'compound fertilizer' and 'micronutrient fertilizer'. Compound fertilizer 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 different interpretations of organic fertilizers. One understanding is naturally occurring or naturally derived fertilizer and the other one refers to the containment of a *significant* mass proportion of the chemical element carbon. The social and the scientific interpretation are not compatible 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 our approach is a scientific one, but also because the AGROVOC thesaurus did not provide any disambiguating hint, we used the reference to carbon to characterize the class 'organic fertilizer' without being able to specify the carbon amount 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 has undergone the process of drying and crushing or powdering. In the very moment they are sold, biofertilizers 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 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 fertilizer contain *significant* mass proportions of organic fertilizers and inorganic fertilizers without clear proportions that would allow the specification of necessary and sufficient membership conditions. Liquid fertilizers and liquid gas fertilizers refer to specific aggregate states at the time of applying these fertilizer types. Slow release fertilizers refer to the characteristic to release plant nutrients slowly without clear boundaries. Fertilizer pesticide combinations also contain *significant* amounts of pesticides.

We also rejected some classes as subtypes of fertilizers, namely 'potting compost' and 'fertilizer combination'. Potting composts do not necessarily contain significant amounts of plant nutrients and fertilizer combinations are fuzzy and impossible to distinguish from other materials.

Fertilizer-related classes

Various classes are closely related to the fertilizer classes because they are fundamental for expressing the membership conditions of the fertilizer classes. A first group of these classes are 'plant nutrient', 'plant micronutrient', 'primary plant nutrient' and 'secondary plant nutrient'. The members of these classes are characterized by their ability to be 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 [48]. E.g., we introduced a class 'plant nutrient 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. The plant nutrient uptake process takes place in 'plants' and has 'plants' as well as 'plant nutrients' as participants.

Overview of the resulting class hierarchy

Figure 5 gives an overview of the fertilizer-related class hierarchy that results from our first characterization of the various fertilizer types. Except rejections of some classes at the bottom end the hierarchy has not changed much in comparison to the original thesaurus hierarchy.

Appendix 4 lists tables that 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 if they are also sufficient conditions expressing a definition for the class. Nevertheless, the tables also contain further information that relates to the results of the alignment process discussed in the next step.

Discussion

The identification of membership conditions, which underlies all subsequent steps, turned out be the crux of formal ontology. A first observation from our re-engineering case study is that the identification of necessary conditions can be greatly facilitated by natural language definitions as a basis. A second observation is that identifying membership conditions stimulates thinking about what the concepts in the thesaurus actually mean in reality and whether the class hierarchy of generic relationship that was adopted from the thesaurus is free of contradictions and consistently narrows the extension.

material
- 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

~~strikethrough~~ – rejected from the ontology as fertilizer type for not fulfilling membership conditions of 'fertilizer' ('potting compost') or for having very vague membership conditions ('fertilizer combination')

*specificity of regulations not sufficient for defining the fertilizer type based on their type-defining chemical component

Figure 5. Fertilizer class hierarchy based on membership conditions extracted from NL definitions; grouped by the possibility to specify them by their containment of nutrients or other chemical elements or molecules

We faced terms in natural language that refer to things in different states. For example, 'composts' may refer to the compost piles before and after their degradation through microorganisms and 'biofertilizers' may refer to organisms as they are sold as product as well as to their state after they have been applied to the field and bound or solubilized plant nutrients. In the ontology it was only one of these states that actually matched our definition of 'fertilizers'. In some cases like 'potting composts' 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 'fertilizers'. Such issues raise the question whether we have to improve the membership conditions that we specified for 'fertilizers' 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 fertilizers' and to decide what to count as 'plant micronutrient'. Overall, the identification of membership conditions clearly faces one with the ambiguities that inhere in a thesaurus and language in general.

Another difficulty that we faced was 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 3 made us revise and rethink our membership conditions. For example, we wondered whether 'composts' are the outcome of the same decomposition process as 'guano' fertilizers or 'green manures'.

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 also connected with a tremendous time effort. It is also the reason, why we emphasize deriving a *semantically adequate* ontology in our method.

3.4. Step 4: Alignment to a top-level ontology and formal relations

Purpose

Alignment is an important step in establishing connections in the ontology that are required for non-trivial logic-based reasoning using reasoning algorithms. Alignment thus facilitates checking the consistency of the ontology, in particular with respect to the is-a hierarchy and the absence of contradiction-free membership conditions. Moreover, alignment may allow further inferences such as the one for new class subsumptions.

Actions to be taken

The activities in this step follow two goals: (1) connecting all class hierarchies to a common top-level ontology and (2) expressing all membership conditions gathered in the previous step through a common set of formally well-defined relationships. To be most effective for the reasoner, the alignment to a top-level ontology is not only done for the classes in the ontology that is developed, but also for classes that are referenced by membership conditions and may be located in external ontologies or may have been newly introduced. The result of this step is an ontology that is tightly integrated with not only the top-level ontology, but also with other ontologies, the classes of which are referenced in the membership conditions.

This step requires two different kinds of activities. First, we need to ensure that all relationships and classes that are necessary to express the membership conditions of the classes in the thesaurus are available. Three activities can be distinguished for this purpose:

a. Choice of an existing top-level ontology and formal relations
b. Choice of existing domain-specific ontologies
c. Amendment of the developed ontology or the external ontologies

Second, all of the ontologies chosen have to be aligned to the top-level ontology, which comprises the following steps:

d. Alignment of the developed ontology to the top-level ontology
e. Alignment of the referenced domain-specific ontologies to the top-level ontology
f. Alignment of the newly introduced classes to the top-level ontology

As a result, all classes in any of the ontologies are subsumed under some class of the chosen top-level ontology. The following subsections will detail the activities in the indicated sequence.

3.4.1. Choice of an existing top-level ontology and formal relations

The choice of an existing top-level ontology is the most fundamental step. It involves getting an overview of existing top-level ontologies and then making a choice between them. To our knowledge there exists neither a registry where all top-level ontologies are listed nor are there guidelines for the choice between top-level ontologies. Some of the commonly cited top-level ontologies include the Descriptive Ontology for Linguistic and Cognitive Engineer-

ing/DOLCE [49], [50], the Basic Formal Ontology/BFO [51], [52], the General Formal Ontology/GFO [53], [54] or the upper levels of CyC [55]–[57]1. Further top-level ontologies are the Suggested Upper Merged Ontology/SUMO [61], [62] or Yet Another More Advanced Top-level Ontology/YAMATO [63], [64]. Borgo and Vieu [58, Sec. 2] give a brief introduction to most of these ontologies. These top-level ontologies are generally published in OWL.

A fixed set of formally defined relationships (object properties in OWL) should be adopted, such as the Relation Ontology [65], [66] or the relationships defined in BioTop [67]. 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 relation, which is a pre-defined part of the Web Ontology Language (OWL).

Application of the step to the fertilizer ontology

While most top-level ontologies are domain-independent, there are also so-called upper-level domain ontologies that describe general kinds of certain domains. Since fertilizers are in the field of biochemistry, we decided to use BioTop [67], [68], an upper-domain ontology for the life sciences. BioTop is particularly suited for our purposes, because it provides (1) a fine-grained distinctions of material entities, (2) a comprehensive set of formally defined relationships [69], 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.

3.4.2. Choice of existing domain-specific ontologies

Top-level ontologies may contain classes, relationships and other entities that are useful for the expression of membership conditions (e.g., to express that portions of agricultural fertilizers are material objects). Obviously, the classes in top-level ontologies are not sufficient for describing the membership conditions of any domain-specific class.

One way to supplement the top-level ontology is by re-using 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 [70] or BioPortal [71]. There are also efforts to build up ontology registries [72] and to develop metadata schemes for such registries [73]. Nevertheless, the current situation to gain an overview of existing ontologies is still far from perfect.

Application of the step to the fertilizer ontology

Since our formal specifications frequently refer to chemical entities, we adopted ChEBI [74], [75], an ontology from the chemistry domain, the major feature of which is the completeness and the hierarchical organization of the chemical elements, molecules and other entities that it models. A disadvantage of ChEBI is that it does not give explicit membership conditions (as of March 2012). It was practical for us that an OWL version of ChEBI is available.

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 multihierarchies 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. This may be said a weakness of Protégé, because classes from a hierarchical path should not be deleted without user interventions, if they also belong to other hierarchical paths.

In principle, we may have been able to adopt further ontologies than ChEBI in order to express membership of fertilizer types. Nevertheless, searching for ontologies and assessing the usefulness of their classes can be time-consuming. Because our main interest was to illustrate the process of choosing and aligning external ontologies, we limited ourselves to ChEBI.

3.4.3. "Amendment" of the external ontologies

If the classes, relationships and other entities that are necessary to express the desired membership conditions are not found in existing ontologies, they

[1] It should be noted that the upper-level hierarchy of CyC is not considered a proper foundational ontology [58, Sec. 2], but rather a result of many historically explicable turns and twists [55], [59] and subject of comprehensive critique [60, Sec. 1].

have to be newly introduced in a way that makes them appear like an amendment to the external ontologies. Newly created classes should, of course, not duplicate what is already contained in one of the imported ontologies. However, the introduction of new classes is unavoidable if a new domain is to be described, whereas introducing new relationships should be avoided and should be seen as the last resort, as idiosyncratic relationships are a main obstacle for interoperability. In many cases, the urge to introduce new relationships is due to an insufficient ontological analysis. Proliferating relationships in OWL can also severely impede the performance of the reasoning algorithms.

When introducing new classes, a decision has to be made, whether these shall be specified with membership conditions. On the one hand, the membership conditions are a valuable basis for checking the consistency of the ontology and to infer class subsumptions. On the other hand, it entails the same effort that is undergone for the thesaurus concepts to be re-engineered. Further, the membership conditions will, in turn, refer to other classes and so forth. We recommend to specify membership conditions for classes that are at heart of the modeled domain only, and leave 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.

Application of the step to the fertilizer ontology

The specifications of the various fertilizer types required the introduction of the classes that are listed in table 2. Only the classes that are central to the fertilizer domain were specified with membership conditions. We did not introduce new relationships in the current development step, because BioTop (which we adopted as our top-level ontology) already contained all the relationships needed (i.e., in the Protégé lingo, all necessary object properties). We will explain in step 6, why we introduced new relationships for data properties, which are not contained in BioTop at all.

3.4.4. Alignment of the thesaurus to the top-level ontology

Aligning a thesaurus to a top-level ontology and a corresponding set of formal relationships includes:

i. organizing all thesaurus concepts into an is-a hierarchy of ontology classes;

Table 2. Classes added to the adopted ontologies

Class	Ontology to which added	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

ii. asserting the top-level thesaurus concepts to be equivalent to appropriate classes of the chosen top-level ontology or subclasses of them; and

iii. expressing membership conditions through the adopted set of formal relationships (like 'has-abstract-part' or 'grain-of').

Organizing all thesaurus concepts into an is-a hierarchy of ontology classes (point i.) is of considerable importance, since all membership conditions and other formal specifications of superordinate classes (e.g. disjointness from other classes) are inherited through is-a relations. They allow for most economic specifications of membership conditions for a class. The generic relationships in a thesaurus are prima facie candidates for becoming is-a relationships in an ontology. Since they may be mixed with hierarchical whole-part relationships in a thesaurus, organizing thesaurus concepts into an is-a hierarchy may imply re-combining fragments of the thesaurus that are not related by properly applied generic relations. This, in turn, may require introducing new classes to connect these fragments. The is-a relationships resulting from the alignment to a top-level ontology are still subject of assessment in step 5 of our method.

Special consideration should be given to poly-hierarchies. As described in [76] and [77, Sec. 1.8], ontologically "correct" poly-hierarchies in the sense that no conflicting membership conditions are inherited from the various hierarchical paths are rare in practice (also called "multiple inheritance problem" or "diamond problem"). Frequently, the existence of poly-hierarchies indicates mistakes in the is-a hierarchy. However, there are ontologically correct poly-hierarchies [78]. It is these hierarchies that are addressed in step 7.

Secondly, we have to assert that the top-level thesaurus concepts are equivalent to classes of the chosen top-level ontology or subclasses of them (point ii.). This requires checking whether the membership conditions of the classes in the top-level ontology apply to all the respectively subsumed thesaurus concepts. This step is interdependent both with the previous point and the next step (step 5) of our method.

Expressing the membership conditions through the formal relationships (point iii.) refers to selecting relationships that semantically express the membership conditions determined in step 3. The selection process is tightly related to the adoption and amendment of formal relationships (the previous activities of the current step) and one has to respect the formal properties of these relationships such as their domain, range, transitivity, disjointness, inverse implication or reflexivity [35, Sec. 9]. 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.

Application of the step to the fertilizer ontology

Since the AGROVOC concepts concerned with agricultural fertilizers are ordered hierarchically by the generic relationship only, we were able to adopt these as is-a relationships in our fertilizer ontology, albeit they are subject of further validation. In connection with its formal specification (see next step), we defined the class 'fertilizer' to be a subclass of the BioTop class 'compound of collective material entities'. Collective material entities are amounts of molecules. Compounds of collective material entities represent the combination of several "pure" materials [67, p. 2008]. We must declare 'fertilizer' to be such compound since there is hardly any pure fertilizer material in real-life environments. Instead, there will always be contained other substances—at least in minimal amounts—and we want to include these under the material we specify here.

The adoption of the relationships from BioTop to express the membership conditions identified in step 3 took considerable time, particularly for familiarizing ourselves with the relationships. The natural language formulations of the membership conditions in the previous section and presented in appendix 4 has already been adjusted to the formulations of relationships in BioTop so that the formalization in step 5 can be easily followed.

3.4.5. Alignment of the referenced domain-specific ontologies to the top-level ontology

Aligning external ontologies to the top-level ontology is done by aligning the top-level classes of these ontologies via the subclass-of or the equivalent-to relationships to adequate classes of the top-level ontology (equivalent to step (ii) of the previous thesaurus alignment activity). Ideally, this has been done by the developers of the adopted ontologies already, but this cannot be taken for granted. In such cases it may be desirable to make at least minimal alignments in order to obtain useful reasoning results.

Application of the step to the fertilizer ontology

We selectively aligned some of the most fundamental classes from ChEBI to the chosen top-level ontology BioTop. To our knowledge such alignments have not been done elsewhere. The first three entries in table 3 show the classes that were aligned (implicitly aligning the subordinate classes) indicating the alignments 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 3. The respective entries in the 2nd column indicate the newly asserted membership conditions.

Table 3. Amendments of necessary membership conditions to existing ChEBI classes

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

3.4.6. Alignment of the newly introduced classes to the top-level ontology

The newly introduced classes should also 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 one adopts the membership conditions from the superordinate classes. In cases of doubt, the class in question should be subsumed under a more general class.

Application of the step to the fertilizer ontology

Table 2 indicates the ontology (BioTop or ChEBI) to which classes 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 alignment that we felt very confident about and, for this reason, aligned to a quite general class at times.

3.4.7. Discussion

The alignment step led to a state where our fertilizer ontology, the top-level ontology (BioTop), other domain-specific ontologies (ChEBI) and their respective amendments are densely interlinked through membership conditions. While some authors have doubts about the usefulness of top-level ontologies [79, p. 12], our experience in this step was that they had an important guiding function by asking us to make categorial distinctions and decisions. BioTop also presented itself as a bundle of highly helpful micro-theories about ontological problems, for example the differentiation of part-of relationships or the distinction between dependent and independent entities. Thus, BioTop with its categorial distinctions and its set of formally defined relationships took many decisions from us, potentially avoided wrong conclusions and mistakes that would otherwise be typical for ad-hoc approaches to ontology development. Instead, we could concentrate on our development task and did not spend time building our own "worldview" out of different scientific papers or other publications. In this sense, BioTop as a top-level ontology formed a counterpart to the ISO thesaurus standard with respect to providing the most fundamental relationships. The breadth of ontological relationships is, of course, far wider than the one in thesauri, which naturally makes ontology development more complex and thus more time-consuming and costly.

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 [49], [50] or BFO [51], [52] and the Relation Ontology [65], [66]), 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.

Alignment also has its price. The alignment was connected with a considerable effort, in particular with respect to understanding the adopted ontologies and relationships. We were faced with various difficult decisions such as which top-level ontology to adopt, when to adopt classes from other domain-specific ontologies (as opposed to defining classes) or which of the classes that we introduced ourselves we should specify in detail (as opposed to "only" subsuming them under some existing classes). There is barely any guidance in the literature for performing these tasks.

It would have been a tremendous advantage, if ChEBI had been more mature in terms of the membership conditions specified for its classes. It would have saved us tremendous time and spared us to deal with amendments of ChEBI. Amendments and also alignments by people other than the developers of an ontology are always connected with great uncertainties, because they are often not fully familiar with the subject area. Finally, the multihierarchy in ChEBI made its trimming for improving the reasoning performance difficult. It is better to avoid and remove poly-hierarchies as we do it in step 7.

We also faced problems with regards to BioTop. It was not absolutely clear to us, if we should model a fertilizer *disposition*, a fertilizer *function* or even a fertilizer *role*. These distinctions need better clarification and guidance. This problem also applies to BFO [51], [52].

3.5. Step 5: Formal specification of membership conditions

Purpose

Membership conditions, alignments and further adaptations that result from the previous two steps are implemented in the chosen formal language in this step so that a reasoner can interpret and check them. Thus, this step continues and alters the formalization started in step 2.

Actions to be taken

Since it may be unusual to formalize the ontology directly in the chosen description language, the formal specification of classes/membership conditions can be subdivided in:

a. Choice of an ontology editor and reasoning algorithm
b. Formalizing the class specifications

c. Adding natural language definitions and comments as class annotations
d. Consistency check and inference of class hierarchy

There are some reviews of ontology editors [80], [81], [82, Ch. 2], and particularly OWL is supported by a growing number of tools. In an online survey, Protégé has been identified as the most popular tool for ontology development [81] (though it has to be taken into account that this survey does not prove the statistical significance of the results). There is also a variety of reasoning algorithms available. The choice of a reasoner depends on various factors, among which the performance certainly is one of the most important ones [83].

The formal specification of classes is realized by adding the necessary membership conditions identified in step 3 as anonymous *superclasses* using the subclass axiom. It is then called a *primitive class* [84, Sec. 4.10]. The specification of a class through necessary and sufficient conditions is realized by adding them as anonymous *equivalent classes* using the equivalent class axiom. It is then called a *defined class* [84, Sec. 4.10]. The terms "defined class" and "primitive class" became more widely used, for example in the popular ontology editor Protégé [85].

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. The consistency check is an automated procedure based on which the reasoning algorithm should point at eventually detected contradictions such as conflicting membership conditions or class subsumption under two disjoint classes. A reasoner can also automatically infer new subsumptions, equivalences or other axioms, if they are entailed logically by the (manually) asserted ontology.

The formal specification of classes by membership conditions is also the step where guidelines for the correct and complete use of OWL [84], [86], [87] or logical ontology design patterns for circumventing expressivity problems of a formal languages [88] should be applied. Following the guidelines may also imply defining additional axioms such as the disjointness of classes or the transitivity of relationships. Further, it is advisable to adopt RFC 3986 [89] or other conventions for the names of the entities (the identifiers, called URIs/IRIs in OWL) in the ontology description.

Based on the formal expression of the membership conditions, there arises the need to distinguish between the asserted ontology on the one hand and the inferred ontology on the other hand. The asserted ontology contains asserted statements only, while the inferred ontology also comprises the inferred statements. When speaking about "the" ontology, the reference is generally to the asserted ontology.

Application of the step to the fertilizer ontology

We chose the Protégé-OWL editor [85] to formalize the ontology in OWL. In terms of formalizing the class specifications, the natural language formulations of membership conditions—concisely summarized in appendix 4—translate relatively easily into OWL class expressions. Only some classes like 'fish manure' and 'guano' have complex membership conditions and thus also complex 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 case of classes that are defined with necessary and sufficient conditions, the 'equivalentTo' axiom applies and the subclass condition becomes 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 lay out a straightforward to express proportions. Simply adding annotation 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 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. Data properties in OWL (the DatatMinCardinality axiom), e.g.

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

are not preoccupied with what their values express. Nevertheless, this condition has to *amend* the condition

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

rather than substituting it. In this solution, the quantity remains machine-readable, but there has to be created a hierarchy of data properties that parallels the hierarchy of chemical elements with the disposition of acting as plant nutrients. 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. One further possibility to express minimum quantities is using additional tools like databases, but this is outside ontological modelling and not in the scope of our case study.

Another general problem with using object properties and data properties in OWL and Protégé is that the quantities cannot be expressed in percentages, but using natural numbers only. This problem can be circumvented 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).

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 [90] which is available as an embedded plug-in for the Protégé-OWL editor.

The reasoning process revealed various initial modeling mistakes that are similar to those described in literature [84], [86], [87] and that we subsequently resolved. Moreover, it 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 decided to check manually, if there can be inferred valid subsumptions considering the minimum plant nutrient levels. For this purpose we kept the concerned classes as defined classes even without the data property-related conditions, and sorted out wrongly inferred subsumptions manually (stroke through in figure 6). 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 6). 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 assuming that the plants used as green manure undergo the same degradation process as other material that is usually referred to as "compost".

Disucssion

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).

At this point it is also worth highlighting some important *limitations of OWL*, which are often discussed as *expressivity limitations*. While there are elaborations of very specific limitations of OWL [91], analyses of fundamental problems using OWL like the one by Stevens et al. [92] are rather rare. Here we

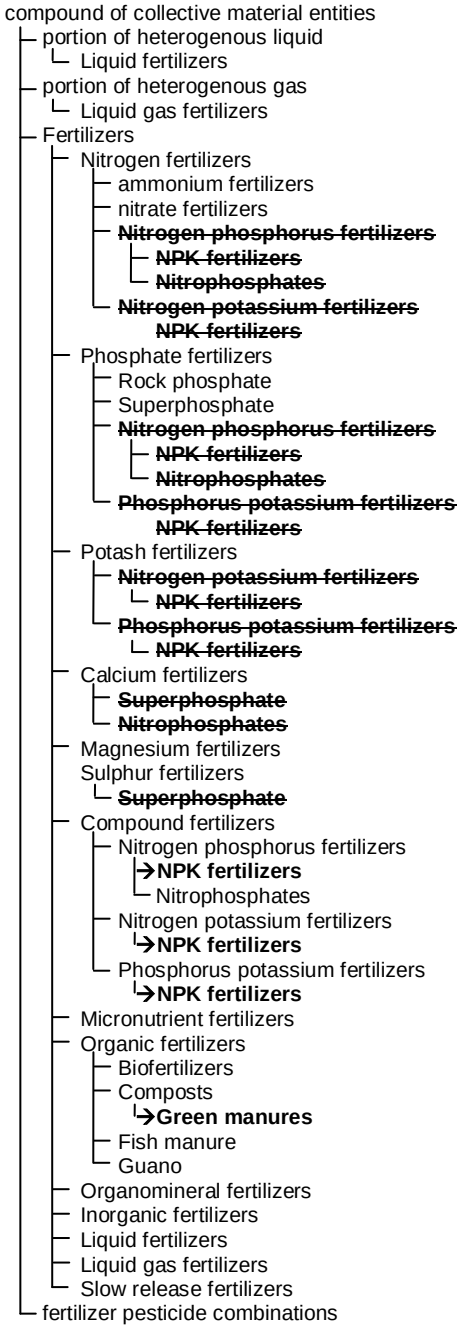

Figure 6. Inferred fertilizer class hierarchy after alignment

want to list some rather macroscopic problems when using OWL and possibly description logics general. Some of them were also described by Saeed [43, Ch. 10] in the context of using formal logics for describing the meaning of natural language statements:

- OWL is limited to countable quantifiers (*all*, *some*, *min* x, *max* y). There are no proportional quantifiers (e.g. most, nearly) and statements like "snow is mostly white" are not possible.
- Unlike some forms of modal logic, OWL has no primitives that could express the modality of a statement, i.e., which qualify a statement through modals such as *usually*, X *thinks that*/*believes*/*is certain that*/*supposes*, it is *likely*/*forbidden*/*desired* 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. There cannot be indicated *when* or *under what circumstances* a certain statement was given or when it will be true.
- 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., statement like "some fertilizers pollute soil" are not possible, but only "all fertilizers pollute soil".

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

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

Purpose

In this step, the labels of classes and other entities are adjusted according to a convention. This improves both readability and understandability 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 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 [93]. Nevertheless, common practices have been summarized [94] and it ought to be checked if 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 one naming style consistently for all classes.

It should be noted that the labeling described here does not concern the name (URI/IRI) of the classes or properties as specified in RFC 3986 [89]. We neither discuss the options for retaining synonym sets from the source thesaurus using the labeling provisions of the respective ontology language, because it does not concern the structure of ontologies that we focus on. Nevertheless, the integration of synonymous may be useful for some applications of ontologies.

Application of the step to the fertilizer ontology

We adopted common conventions in biomedical ontologies for the class labelling summarized by Schober et al. [94]. 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-labelled to begin with "portion of" to emphasize that we deal with amounts of materials, not with countable objects. The term "fertilizer" was added to the classes labelled "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" often adds an emphasis on the fact 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 only the outcome of specific processes to which the previously mentioned materials are input. In case of 'fish manure' we adopted the commonly used term "fish fertilizer". Appendix 5 provides a complete overview of the labeling changes.

3.7. Step 7: Dissolving poly-hierarchies

Purpose

In order to get an ontology that can easily be maintained, poly-hierarchies should be dissolved in the ontology. This concerns only the semantically correct poly-hierarchies that do not inherit contradictory membership conditions from their superordinate classes. Such incorrect poly-hierarchies should have been removed in step 4 (discussed in subsection 3.4). Dissolving poly-hierarchies is an optional step, since it does not change the semantics of the ontology.

Actions to be taken

Dissolving poly-hierarchies requires a decision as to which one of two or more hierarchical class paths shall be retained, that is, which single direct superclass is to be kept out of several available direct superclasses. The other direct superclasses are "dissolved" in the sense that (a) the restrictions of the classes along the dissolved class paths are added to the specification of the target class and (b) any subsumption of the target class under classes of the dissolved class path is removed from the specification of the target class.

Dissolving poly-hierarchies in the *asserted* ontology in such way is one aspect of the "normalization" method recommended by Rector [76]. Notably, the methodical step never results in any loss of semantic information. The poly-hierarchies can later be automatically restored through automated reasoning, thus becoming part of the *inferred* ontology.

Application of the step to the fertilizer ontology

In the ontology that we have modelled, 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 'liquid fertilizer' here, illustrated in figure 7.

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 7) by adding a membership condition to the specification of the class 'liquid fertilizer' (namely *'bearer of' some*

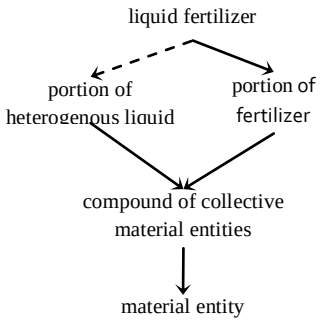

Figure 7. Poly-hierarchy for 'liquid fertilizer' (the dotted arrow indicates the is-a relationship dissolved by us).

('quality located' some 'liquid value region'), which are all classes and relationships in BioTop). Of course, membership conditions that are already part of the 'liquid fertilizer' specification or its superclasses 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:
'liquid fertilizer' EquivalentTo (fertilizer and 'portion of heterogenous liquid')

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

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

Discussion

Dissolving poly-hierarchies is a straightforward step. The decision, whether or not to implement this step, is partially a matter of personal preference. Mono-hierarchies are easier to implement and to maintain, but sometimes it might be intellectually challenging to decide which is-a relation is to be dissolved.

4. 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 subactivities. In this section we will reflect on the method overall, 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 a 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 reality. The steps of our method achieve this quality 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 the classes is expressed through membership conditions (step 3).
(3) Newly created as well as imported classes are aligned to a top-level ontology; and a common set of formal relationships is used (step 4).
(4) The ontology is checked for consistency and the inferences that can be drawn from the asserted ontology (the logically inferred subsumptions or other axioms) have been checked for plausibility (step 5).
(5) The ontology has a rigorous is-a hierarchy in which the intension of classes (the specification of the classes) is becoming 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 intend to define one understanding of a natural language term.

Requirement (5) may not be obvious, but it is based on the adoption of the generic relationship in a thesaurus as is-a hierarchy and its gradual refinement by grounding it on membership conditions (step 3), adopting high-level membership conditions through the alignment to a top-level ontology (step 4) 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 subject of further investigations. The rigorous is-a hierarchy makes ontologies especially apt 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* be assumed to *always* 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 eventually sufficient membership conditions (step 3). Determining minimum proportions of plant nutrients in fertilizers and formalizing these in OWL have literally become studies in their own rights. It took also considerable time to get adjusted to the framework of BioTop and the ChEBI ontology to express the membership conditions using these ontologies (step 4).

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

- The effort with the preparation and checking of 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 3).
- Experience with the chosen top-level ontology and other imported ontologies reduces the alignment effort (step 4).
- Experience in modelling with OWL reduces the effort with the correct formal specification of membership conditions (step 5).
- Optional steps and sub-activities such as adjusting entity labels (step 6) dissolving polyhierarchies (step 7) or 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.

The generality of our method, i.e. its applicability to all existing thesauri, is guaranteed by step 1, which demands the preparation and checking of the thesaurus with respect to the thesaurus standard ISO 25964-1:2011. While we had to deal with various differences and similarities in the case study that were theoretically anticipated in a prior comparative study of relata and relationships in thesauri and ontologies [34], we did not face all these differences in the 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 the natural and life sciences. It is an open question, whether one faces greater problems when applying our method in other domains such as the social sciences. For example, it may be more difficult to define membership conditions for concepts like ‘freedom’ or ‘success’ than for material objects or phenomena that can be analyzed and measured objectively with instruments such as sensors. This does not question the applicability of our re-engineering method as such, but rather questions the usefulness of ontologies described in OWL in specific domains overall. The agricultural domain of the case study may have favored the application of the re-engineering method.

The method used a thesaurus as a starting point for the 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.

5. Relation to existing re-engineering methods

Because we have fully explained our re-engineering method at this point, it is also easier to understand, how our method differs from existing re-engineering methods. In this section we will start with characterizing our method as T-Box re-engineering for which there exist no methods at this point of time. Subsequently, we will introduce commonly applied A-Box re-engineering methods as well as a number of other understandings of ontologies and methods for re-engineering thesauri into ontologies. We will explain that these understandings and methods are unrelated and, in fact, incompatible with our understanding of ontologies and re-engineering.

The basic premise of our re-engineering approach rests on the distinction and purpose of the TBox and ABox in OWL and other description logics. While the TBox “contains intensional knowledge in the form of a terminology and is built through declarations that describe general properties of concepts”, the ABox “contains extensional knowledge—also called assertional knowledge—knowledge that is specific to the individuals of the domain of discourse.” [95, Sec. 1.3]. In other words, the TBox (sometimes called the “vocabulary”) concentrates on the intensional specification of classes using previ-

ously specified relationships while the ABox uses the definitions made in the TBox to describe particular things (individuals) in the real word. The TBox acts thus as a metamodel for the ABox, "a model that consists of statements about models" [96]. We follow Guarino et al. [97] in considering only intensional knowledge (the TBox) to be part of an ontology.

Concepts in thesauri are—with some exceptions—intensional entities that are labelled by general terms, terms that are "predicable, in the same sense, of more than one individual" [39, p. 544]. As figure 8 shows, re-engineering thesauri into ontologies thus means that the majority of the thesaurus content (b) ends up in the TBox (2). Only very few thesaurus concepts, in particular references to instances of the actual world such as the "Mekong River" or "Rocky mountains", end up in the ABox, but are then not considered part of the ontology (TBox). 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 (1)[2]. With "data model" we refer to a model that "determines the logical structure of a database and fundamentally determines in which manner data can be stored, organized, and manipulated" [98], often called database model.

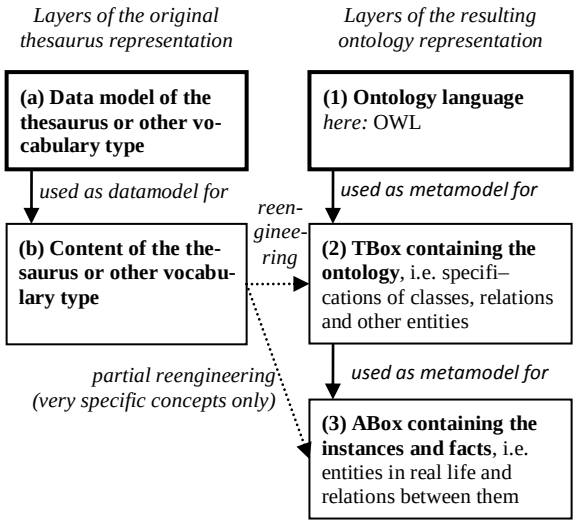

Figure 8. TBox re-engineering process for thesauri and other types of vocabularies

[2] Ontologies described in the TBox are sometimes referred to as formal ontologies in order to contrast them to "ABox ontologies" that tend to be called lightweight ontologies in this context. In this thesis only formal ontologies are considered ontologies while lightweight ontologies are not considered ontologies at all.

This approach is generally referred to as "TBox re-engineering". Our method is the first one that systematically describes such TBox re-engineering. Only very few authors follow this understanding of an ontology when reporting about their efforts of re-engineering specific thesauri. Among these authors are **Hahn** [7] and **Hahn and Schulz** [99], whose recommendations are based on their experience with the UMLS meta-thesaurus. **Wroe et al.** [9] dealt with the Gene Ontology. Table 4 gives an overview of the methodical steps that we could identify in these publications and how they relate to the steps that we presented in our method.

Table 4. Methodical steps for the ontological re-engineering of thesauri identified in literature

Methodical step	Reference backing the step	Corresponding step in our re-engineering method
a) Refinement and completion of formal specifications	Hahn [7], Wroe et al. [9]	Steps 2 and 3
b) Identification and removal of cycles in the is-a hierarchy	Hahn [7]	Step 2
c) Syntactic translation	Hahn [7], Wroe et al. [9]	Step 2
d) Application of a top-level ontology	Hahn [7]	Step 4

Apart from these specific reports, none of which provides a detailed instructive description of steps, there is no method that holistically describes (TBox) re-engineering. The report of the **NeOn project** [12] mentions TBox re-engineering, but in the end refers to some software or algorithm called Scarlet [100] and the use of WordNet. The use of these instruments is not explained. The contribution to TBox re-engineering and thus ontological re-engineering remains unclear.

Although not being a re-engineering method as such, OntoClean [101], [102] is the only method that we consider closely related to our re-engineering method. OntoClean is focused on improving the is-a hierarchy, which is also an implicit result of steps 3, 4, 5, and 7 of our method. Particularly the alignment to a top-level ontology in step 4 may have effects on the is-a hierarchy that are comparable to applying 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. We did not detect any errors in the is-a hierarchy when applying OntoClean and thus did not include OntoClean as a step in our method.

The previously described TBox re-engineering can be contrasted to a re-engineering approach that is often called "ABox re-engineering". The major premise of ABox re-engineering is to avoid structural changes of the thesaurus [12, p. 96], which generally makes the re-engineering easy to automate. The basic principle behind ABox-focussed methods is displayed in figure 9. The modelling primitives of an ontology language (1) are used (instantiated) to describe the data model of a given thesaurus or other vocabulary type (a) in the TBox (2). The data model in the TBox is then regarded as the "ontology" and used (instantiated) to describe the content of a domain-specific thesaurus (b) in the ABox (3). An example of such data model in the TBox is SKOS, an abbreviation for "Simple Knowledge Organization System" [103], which is closely oriented on the thesaurus data model described in ISO 25964-1:2011.

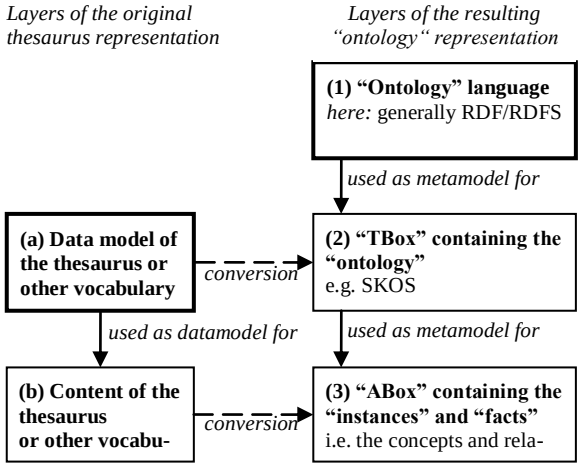

Figure 9. ABox "re-engineering" process for thesauri and other types of vocabularies

The described approach of an ABox re-engineering often goes hand-in-hand with the use of RDF or RDFS that we have already criticized to be an inadequate languages for the description of ontologies in the introduction. The distinction of a TBox and an ABox is neither present nor practically relevant in RDF/RDFS and was displayed here as a contrast to the TBox re-engineering only. OWL would also have to be used in an unconventional way in ABox re-engineering and we also could not observe such attempts in practice.

Examples of ABox re-engineering methods can be found in the PhD thesis of **Villazón-Terrazas** [5], [13] that underlies also the results of the **NeOn project** [12]. The PhD thesis by **van Assem** [4] is an ABox conversion as well and offers the choice between using SKOS [15] and specifying a non-standard data model in the TBox [14]. Van Assem essentially considers differentiating the hierarchical thesaurus relationship into two different relationships—a transitive and a non-transitive one—to be a semantic conversion. These relationships are then defined as a subtype of the subclass relationship in RDFS, although van Assem recognizes himself that this practice is often incorrect.

We consider the ABox re-engineering to be a wrong use of OWL and description logic in general, which misplaces the typical concept of a thesaurus in the ABox. It relates to a widespread understanding of the TBox of an ontology as a data model and not as a specification of membership conditions of entities. Also other authors have criticized the position that the difference between a thesaurus and an ontology is of purely syntactic nature [104, p. 17].

Another group of publications understates re-engineering as simple refinement of the relationships of a thesaurus. The most representative publication in this regards is **Soergel et al.** [6]. This approach underlies various other publications, e.g. **Kawtrakul et al.** [105] or **Sánchez-Alonso and Sicilia** [106] and has been applied to the AGROVOC thesaurus, which was also subject of our re-engineering in section 3. Similar ideas have been presented as the "ontological augmenting of thesaurus relationships" by **Tudhope et al.** [107]. According to Soergel et al. different hierarchical relationships have to be distinguished if, e.g., automated reasoning is to be supported.

Table 5 shows examples of such refinements.

Table 5: Refinement of thesaurus relationships according to Soergel et al. [6]

Sub-relationships of the hierarchical relationship
'Colorado river' instanceOf 'rivers'
'blood' containsSubstance 'blood proteins'
'roots' yieldsPortion 'cuttings'
'Francophone Africa' hasMember 'Benin'
Sub-relationships of the associative relationship
'overgrazing' causes 'desertification'
'plough' instrumentFor 'ploughing'

Our re-engineering confirms that, indeed, thesaurus relationships often need to be refined to become valid relationships in ontologies. Nevertheless, Soergel et al. as well as most of the authors that do not focus on TBox engineering oversee that in an

ontology (1) any relationship from a class A to another class B has *always* the role and logical force of a necessary membership condition for the class A. (2) Relationships involve implicit or (in OWL) explicit quantification, which is relevant for the semantics of relational expressions [108]. Thus, (3) 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. The described characteristics of relationships in ontologies do not necessarily coincide and may even conflict with the rules in thesaurus standards, particularly with respect to the associative relationships in thesauri. Thus, many if not most of the thesaurus relationships have to be *rejected* in an ontology thesaurus.[3]

Other re-engineering methods are even more simple and do not provide deep insights. One example is the method by **Wielinga et al.** [109] who use RDF semantics, which does not distinguish between instances and classes and is thus not of interest here. For **Hepp and de Bruijn** [110] deriving ontologies from hierarchical classifications, thesauri, or inconsistent taxonomies means defining contexts like 'product' or 'service' which can be combined with concepts such as 'TV set' to create categories like 'TV as product' or 'TV as service'. They see this as sufficient for a script-based creation of "meaningful ontology classes", without really saying what purpose this has.

In summary, there are currently no reengineering methods that make use of the semantic capabilities of formal languages like OWL in order to detect logical mistakes and to improve vocabularies. The method that we contributed in this paper is thus unique, although it reflects the way that at least some of the biomedical vocabularies are developed nowadays.

6. Conclusions

We presented a method with seven steps and numerous subactivities for re-engineering thesauri into semantically adequate ontologies using the description logic based OWL format. 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 by making full use of OWL's capabilities 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 logically consistently. These benefits are achieved by imposing a more consistent is-a hierarchy and by removing relationships from thesauri that are not valid in a formal ontology.

[3] These considerations do *not* apply to is-a relationship (the subclass relationship in OWL) and the instance-of relationship (expressed by a class assertion in OWL). With regards to the use of relationships, it should also be noted that ontology work with OWL, description logic and many other deductive logics is *not* interested in any "typical", "usual", or "desired" properties of the concepts. Their inclusion in an ontology generally leads to wrong reasoning results, particularly when integrating different ontologies, and must be considered a wrong use of OWL.

7. Acknowledgements

The research of D.K. has been enabled through the David Hay Memorial Fund and the PORES travel and research grant provided by University of Melbourne, with special thanks to Edmund Kazmierczak and Simon Milton for their support in setting up the research visit. The work of L.J. has been supported by the German Research Foundation (DFG) under the auspices of the GoodOD project.

8. References

[1] F. Baader, I. Horrocks, and U. Sattler, 'Description Logics', in *Handbook on Ontologies*, 2nd ed., S. Staab and R. Studer, Eds. Springer, 2009, pp. 21–43.

[2] E. Simperl, C. Tempich, and Y. Sure, 'Ontocom: A cost estimation model for ontology engineering', in *Proceedings of fifth ISWC*, 2006.

[3] E. Simperl, 'Reusing ontologies on the Semantic Web: A feasibility study', *Data Knowl Eng*, vol. 68, no. 10, pp. 905–925, 2009.

[4] M. van Assem, 'Converting and Integrating Vocabularies for the Semantic Web', Vrije Universiteit, Amsterdam, the Netherlands, 2010.

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

[6] D. Soergel, B. Lauser, A. Liang, F. Fisseha, J. Keizer, and S. Katz, 'Reengineering Thesauri for New Applications: the AGROVOC Example', *J. Digit. Inf.*, vol. 4, no. 4, 2004.

[7] U. Hahn, 'Turning Informal Thesauri Into Formal Ontologies: A Feasibility Study on Biomedical Knowledge re-Use', in *Comparative and Functional Genomics*, 2003, vol. 4, pp. 94–97.

[8] E. Hyvönen, K. Viljanen, J. Tuominen, and K. Seppälä, 'Building a national semantic web ontology and ontology service infrastructure—the FinnONTO approach', in *Proceedings of the 5th European semantic web conference*

ESWC 2008, Tenerife, Spain, June 1-5, 2008, Berlin, Heidelberg, 2008, pp. 95–109.

[9] C. Wroe, R. Stevens, C. A. Goble, and M. Ashburner, 'A methodology to migrate the Gene ontology to a description logic environment using DAML OIL', in *Proceedings of the 8th Pacific Symposium on Biocomputing (PSB)*, Hawaii, 2003, pp. 624–635.

[10] B. Smith and B. Klagges, 'Philosophy and Biomedical Information Systems', in *Applied Ontology. An Introduction*, K. Munn and B. Smith, Eds. ontos verlag, 2009, pp. 21–38.

[11] B. Smith and W. Ceusters, 'Ontological realism: A methodology for coordinated evolution of scientific ontologies', *Appl. Ontol.*, vol. 5, no. 3–4, pp. 139–188, Nov. 2010.

[12] S. Angeletou, H. Lewen, and B. Villazón, 'Methods for re-engineering and evaluation', Open University (OU), Milton Keynes, UK, Deliverable 2.2.4, Integrated Project (IST-2005-027595), version 1.0, Jan. 2010.

[13] B. M. Villazón-Terrazas and A. Gómez-Pérez, 'Reusing and Re-engineering Non-ontological Resources for Building Ontologies', in *Ontology Engineering in a Networked World*, Springer Berlin Heidelberg, 2012, pp. 107–145.

[14] M. van Assem, M. R. Menken, G. Schreiber, J. Wielemaker, and B. Wielinga, 'A Method for Converting Thesauri to RDF/OWL', in *The Semantic Web – ISWC 2004*, vol. 3298, S. A. McIlraith, D. Plexousakis, and F. Harmelen, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 17–31.

[15] M. van Assem, V. Malaisé, A. Miles, and G. Schreiber, 'A Method to Convert Thesauri to SKOS', in *The Semantic Web: Research and Applications*, 2006, pp. 95–109.

[16] RDF Working Group, 'Resource Description Framework (RDF)', 22-Mar-2013. [Online]. Available: http://www.w3.org/RDF/.

[17] D. Brickley and R. V. Guha, Eds., *RDF Vocabulary Description Language 1.0: RDF Schema*. World Wide Web Consortium (W3C), 2004.

[18] T. Berners-Lee, J. Hendler, and O. Lassila, 'The Semantic Web', *Sci. Am.*, vol. 284, no. 5, p. 34, May 2001.

[19] P. Hayes, *RDF Semantics*. World Wide Web Consortium (W3C), 2004.

[20] G. Antoniou and F. van Harmelen, 'Web Ontology Language: OWL', in *Handbook on Ontologies*, 2nd ed., 2009, pp. 91–110.

[21] D. L. McGuinness and F. van Harmelen, Eds., *OWL Web Ontology Language. Overview*. World Wide Web Consortium (W3C), 2004.

[22] W3C OWL Working Group, *OWL 2 Web Ontology Language. Document Overview (Second Edition)*. World Wide Web Consortium (W3C), 2012.

[23] B. Motik, P. F. Patel-Schneider, and B. C. Grau, Eds., *OWL 2 Web Ontology Language. Direct Semantics*. World Wide Web Consortium (W3C), 2009.

[24] F. Baader and W. Nutt, 'Basic Description Logics', in *The description logic handbook: theory, implementation, and applications*, 2nd ed., F. Baader, D. Calvanese, D. L. McGuinness, D. Nardi, and P. F. Patel-Schneider, Eds. Cambridge University Press, 2003, pp. 43–95.

[25] P. F. Patel-Schneider and B. Motik, Eds., *OWL 2 Web Ontology Language. Mapping to RDF Graphs*. World Wide Web Consortium (W3C), 2009.

[26] M. Schneider, Ed., *OWL 2 Web Ontology Language. RDF-Based Semantics*. World Wide Web Consortium (W3C), 2009.

[27] OBO Foundry homepage, 'The Open Biological and Biomedical Ontologies', 25-Oct-2012. [Online]. Available: http://www.obofoundry.org/. [Accessed: 25-Oct-2012].

[28] R. Rocha Souza, D. Tudhope, and M. Barcellos Almeida, 'The KOS spectra: A tentative typology of knowledge organization systems', in *Paragdigms and conceptual systems in knowledge organization: Proceedings of the 11th ISKO International Conference, Rome*, 2010, pp. 122–129.

[29] ANSI/NISO Z39.19-2005, *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*. 2005.

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

[31] Taxonomy Warehouse, 'Taxonomies', 2012. [Online]. Available: http://www.taxonomywarehouse.com/headword_list_new.aspx?vObject=10076&stype=ab. [Accessed: 08-Dec-2012].

[32] AGROVOC, *Agricultural Information Management Standards (AIMS)*, 2012. [Online]. Available: http://aims.fao.org/standards/agrovoc/about. [Accessed: 12-Nov-2012].

[33] D. Kless, L. Jansen, J. Lindenthal, and J. Wiebensohn, 'A method for re-engineering a thesaurus into an ontology', in *Proceedings of the 7th International Conference*, Graz, Austria, 2012, vol. Volume 239, pp. 133–146.

[34] D. Kless, S. Milton, and E. Kazmierczak, 'Relationships and Relata in Ontologies and Thesauri: Differences and Similarities', *Appl. Ontol.*, vol. 7, no. 4, pp. 401–428, Nov. 2012.

[35] B. Motik, P. F. Patel-Schneider, and B. Parsia, Eds., *OWL 2 Web Ontology Language. Structural Specification and Functional-Style Syntax*. World Wide Web Consortium (W3C), 2009.

[36] P. Hitzler, M. Krötzsch, B. Parsia, P. F. Patel-Schneider, and S. Rudolph, Eds., *OWL 2 Web Ontology Language. Primer*. World Wide Web Consortium (W3C), 2009.

[37] K. La Barre, 'Facet analysis', *Annu. Rev. Inf. Sci. Technol.*, vol. 44, no. 1, pp. 243–284, 2010.

[38] A. L. D. Brockmöller, 'Ontological Thesaurus Extension: the AAT example', University of Amsterdam (UvA), Amsterdam, the Netherlands, 2003.

[39] B. A. Brody, 'Logical terms, glossary of', *Encyclopedia of Philosophy*, vol. 5, 10 vols. Macmillan Reference, USA, pp. 533–560, republished without changes in 2005-1967.

[40] A. Isaac and E. Summers, Eds., *SKOS Simple Knowledge Organization System. Primer*. World Wide Web Consortium (W3C), 2009.

[41] H. Putnam, 'It ain't necessarily so', *J. Philos.*, vol. 59, no. 22, pp. 658–671, 1962.

[42] S. A. Kripke, *Naming and necessity*. Oxford: Blackwell, 1980.

[43] J. I. Saeed, *Semantics*, 3rd ed. Wiley-Blackwell, 2009.

[44] V. Gowariker, V. N. Krishnamurthy, S. Gowariker, M. Dhanorkar, and K. Paranjape, *The Fertilizer Encyclopedia*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2008.

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

[46] J. Paavola, 'Resources', *International Encyclopedia of the Social Sciences*. Encyclopedia.com, 2008.

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

[48] J. Röhl and L. Jansen, 'Representing dispositions', *J. Biomed. Semant.*, vol. 2, no. Suppl 4, p. S4, Aug. 2011.

[49] A. Gangemi, N. Guarino, C. Masolo, A. Oltramari, and L. Schneider, 'Sweetening ontologies with DOLCE', *Knowl. Eng. Knowl. Manag. Ontol. Semantic Web*, pp. 223–233, 2002.

[50] S. Borgo and C. Masolo, 'Ontological Foundations of DOLCE', in *Theory and Applications of Ontology: Computer Applications*, R. Poli, M. Healy, and A. Kameas, Eds. Dordrecht: Springer Netherlands, 2010, pp. 279–295.

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

[52] H. Stenzhorn, 'Homepage', *Basic Formal Ontology (BFO)*, 13-Aug-2012. [Online]. Available: http://www.ifomis.org/bfo/. [Accessed: 25-Oct-2012].

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

[54] GFO homepage, 'General Formal Ontology (GFO)', 2010. [Online]. Available: http://www.onto-med.de/ontologies/gfo/. [Accessed: 25-Oct-2012].

[55] E. Bertino, B. Catania, and G. P. Zarri, 'The Cyc project', in *Intelligent database systems*, Addison-Wesley Professional, 2001, pp. 275–316.

[56] D. Foxvog, 'Cyc', in *Theory and Applications of Ontology: Computer Applications*, R. Poli, M. Healy, and A. Kameas, Eds. Dordrecht: Springer Netherlands, 2010, pp. 259–278.

[57] Cycorp, 'Diagram of the OpenCyc Upper Ontology'. 27-Mar-2002.

[58] S. Borgo and L. Vieu, 'Artefacts in Formal Ontology', in *Philosophy of Technology and Engineering Sciences*, Amsterdam: North-Holland, 2009, pp. 273–307.

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

[60] L. Jansen, 'Categories: The Top-Level Ontology', in *Applied Ontology. An Introduction*, K. Munn and B. Smith, Eds. ontos verlag, 2009, pp. 173–196.

[61] I. Niles and A. Pease, 'Towards a standard upper ontology', in *Proceedings of the international conference on Formal Ontology in Information Systems-Volume 2001*, 2001, pp. 2–9.

[62] SUMO homepage, *The Suggested Upper Merged Ontology (SUMO)*, 18-Jul-2012. [Online]. Available: http://www.ontologyportal.org/. [Accessed: 25-Oct-2012].

[63] R. Mizoguchi, 'YAMATO: Yet Another More Advanced Top-level Ontology', in *Proceedings of the Sixth Australasian Ontology Workshop*, 2010, pp. 1–16.

[64] YAMATO homepage, 'YAMATO: Yet Another More Advanced Top-level Ontology', 15-Dec-2010. [Online]. Available: http://www.ei.sanken.osaka-u.ac.jp/hozo/onto_library/upperOnto.htm. [Accessed: 25-Oct-2012].

[65] B. Smith, W. Ceusters, B. Klagges, J. Köhler, A. Kumar, J. Lomax, C. Mungall, F. Neuhaus, A. Rector, and C. Rosse, 'Relations in biomedical ontologies', *Genome Biol.*, vol. 6, no. 5, p. R46, Apr. 2005.

[66] OBO relations homepage, [Online]. Available: http://code.google.com/p/obo-relations/. [Accessed: 25-Oct-2012].

[67] E. Beisswanger, S. Schulz, H. Stenzhorn, and U. Hahn, 'BioTop: An upper domain ontology for the life sciencesA description of its current structure, contents and interfaces to OBO ontologies', *Appl. Ontol.*, vol. 3, no. 4, pp. 205–212, 2008.

[68] S. Schulz, 'BioTop - A Top-Domain Ontology for the Life Sciences', Jan-2012. [Online]. Available: http://www.imbi.uni-freiburg.de/ontology/biotop/. [Accessed: 25-Oct-2012].

[69] S. Schulz and U. Hahn, 'Towards the ontological foundations of symbolic biological theories', *Artif Intell Med*, vol. 39, no. 3, pp. 237–250, Mar. 2007.

[70] OBO Download Matrix, 13-Jun-2012. [Online]. Available: http://www.berkeleybop.org/ontologies/. [Accessed: 25-Oct-2012].

[71] P. L. Whetzel, N. F. Noy, N. H. Shah, P. R. Alexander, C. Nyulas, T. Tudorache, and M. A. Musen, '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.*, vol. 39, no. Web Server issue, pp. W541–W545, Jul. 2011.

[72] M. d' Aquin and N. F. Noy, 'Where to publish and find ontologies? A survey of ontology libraries', *Web Semant. Sci. Serv. Agents World Wide Web*, 2011.

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

[74] K. Degtyarenko, P. de Matos, M. Ennis, J. Hastings, M. Zbinden, A. McNaught, R. Alcantara, M. Darsow, M. Guedj, and M. Ashburner, 'ChEBI: a database and ontology for chemical entities of biological interest', *Nucleic Acids Res.*, vol. 36, no. Database, pp. D344–D350, Dec. 2007.

[75] ChEBI homepage, 'Chemical Entities of Biological Interest (ChEBI)', 2012. [Online]. Available: http://www.ebi.ac.uk/chebi/. [Accessed: 02-Nov-2012].

[76] A. Rector, 'Modularisation of domain ontologies implemented in description logics and related formalisms including OWL', in *Proceedings of the 2nd international conference on Knowledge capture*, 2003, pp. 121–128.

[77] G. H. Merrill, 'Ontological realism: Methodology or misdirection?', *Appl. Ontol.*, vol. 5, no. 2, pp. 79–108, Jun. 2010.

[78] I. Johansson, 'Four Kinds of "Is_A" Relation', in *Applied Ontology. An Introduction*, K. Munn and B. Smith, Eds. ontos verlag, 2009, pp. 235–254.

[79] D. B. Lenat, 'Applied ontology issues', *Appl. Ontol.*, vol. 1, pp. 9–12, Jan. 2005.

[80] M. Denny, 'Ontology tools survey, revisited', *XML.com*, 2004.

[81] M. R. Khondoker and P. Mueller, 'Comparing Ontology Development Tools Based on an Online Survey', 2010.

[82] A. Gómez-Pérez, 'A survey on ontology tools', Vrije Universiteit Amsterdam (VU), Amsterdam, the Netherlands, Deliverable Deliverable 1.3, IST-2000-29243, May 2002.

[83] K. Dentler, R. Cornet, A. ten Teije, and N. de Keizer, 'Comparison of reasoners for large ontologies in the OWL 2 EL profile', *Semantic Web*, vol. 2, no. 2, pp. 71–87, 2011.

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

[85] Protégé-OWL editor, 'What is Protégé-OWL?', 2012. [Online]. Available: http://protege.stanford.edu/overview/protege-owl.html. [Accessed: 25-Oct-2012].

[86] A. Borgida and R. J. Brachman, 'Conceptual Modeling with Description Logics', in *The description logic handbook: theory, implementation, and applications*, 2nd ed., F. Baader, D. Calvanese, D. L. McGuinness, D. Nardi, and P. F. Patel-Schneider, Eds. Cambridge University Press, 2003, pp. 349–372.

[87] A. Rector, N. Drummond, M. Horridge, J. Rogers, H. Knublauch, R. Stevens, H. Wang, and C. Wroe, 'OWL Pizzas: Practical Experience of Teaching OWL-DL: Common Errors & Common Patterns', in *Engineering Knowledge in the Age of the Semantic Web*, vol. 3257, E. Motta, N. R. Shadbolt, A. Stutt, and N. Gibbins, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 63–81.

[88] V. Presutti, A. Gangemi, D. Stefano, G. A. de Cea, M. C. Suárez-Figueroa, E. Montiel-Ponsoda, and M. Poveda, 'A Library of Ontology Design Patterns', Consiglio Nazionale delle Ricerche (CNR), Roma-Lazio Italy, Deliverable 2.5.1, Integrated Project (IST-2005-027595), version 1.2, Feb. 2008.

[89] RFC 3986, 'Internationalized Resource Identifiers (IRIs)', Request for Comments, Jan. 2005.

[90] HermiT OWL Reasoner, [Online]. Available: http://hermit-reasoner.com/. [Accessed: 14-Nov-2012].

[91] B. C. Grau, I. Horrocks, B. Motik, B. Parsia, P. F. Patel-Schneider, and U. Sattler, 'OWL 2: The next step for OWL', *Web Semant. Sci. Serv. Agents World Wide Web*, vol. 6, no. 4, pp. 309–322, Nov. 2008.

[92] R. Stevens, M. Egaña Aranguren, K. Wolstencroft, U. Sattler, N. Drummond, M. Horridge, and A. Rector, 'Using OWL to model biological knowledge', *Int. J. Hum.-Comput. Stud.*, vol. 65, no. 7, pp. 583–594, Jul. 2007.

[93] V. Svátek, O. Šváb-Zamazal, and V. Presutti, 'Ontology naming pattern sauce for (human and computer) gourmets', in *Workshop on Ontology Patterns at ISWC*, 2009, vol. 9.

[94] D. Schober, B. Smith, S. Lewis, W. Kuśnierczyk, J. Lomax, C. Mungall, C. Taylor, P. Rocca-Serra, and S. A. Sansone, 'Survey-based naming conventions for use in OBO Foundry ontology development', *BMC Bioinformatics*, vol. 10, no. 1, p. 125, 2009.

[95] D. Nardi and R. J. Brachman, 'An Introduction to Description Logics', in *The description logic handbook: theory, implementation, and applications*, 2nd ed., F. Baader, D. Calvanese, D. L. McGuinness, D. Nardi, and P. F. Patel-Schneider, Eds. Cambridge University Press, 2003, pp. 1–40.

[96] M. A. Jeusfeld, 'Metamodel', *Encyclopedia of Database Systems*. Springer, pp. 1727–1730, 2009.

[97] N. Guarino, D. Oberle, and S. Staab, 'What is an Ontology?', in *Handbook on Ontologies*, 2nd ed., S. Staab and R. Studer, Eds. Springer, 2009, pp. 1–17.

[98] Database model, *Wikipedia, the free encyclopedia*. 16-Mar-2013.

[99] U. Hahn and S. Schulz, 'Ontology engineering by thesaurus re-engineering', in *Information Modelling and Knowledge Bases Xiii*, H. Kangassalo, Ed. IOS Press, 2002.

[100] Scarlet, 2010. [Online]. Available: http://scarlet.open.ac.uk/. [Accessed: 25-Oct-2012].

[101] N. Guarino and C. Welty, 'An overview of OntoClean', *Handb. Ontol.*, pp. 151–159, 2004.

[102] N. Guarino and C. Welty, 'An Overview of OntoClean', in *Handbook on Ontologies*, 2nd ed., 2009, pp. 201–220.

[103] A. Isaac, 'Homepage', *SKOS: Simple Knowledge Organization System*, 01-Jan-2012. [Online]. Available: http://www.niso.org/schemas/iso25964/. [Accessed: 17-Sep-2012].

[104] W. Ceusters, B. Smith, and L. Goldberg, 'A terminological and ontological analysis of the NCI Thesaurus', *Methods Inf. Med.*, vol. 44, no. 4, p. 498, 2005.

[105] A. Kawtrakul, A. Imsombut, A. Thunkijjanukit, D. Soergel, A. Liang, M. Sini, G. Johannsen, and J. Keizer, 'Automatic term relationship cleaning and refinement for AGROVOC', in *Workshop on The Sixth Agricultural Ontology Service*, 2005, pp. 247–260.

[106] S. Sánchez-Alonso and M. A. Sicilia, 'Using an AGROVOC-based ontology for the description of learning resources on organic agriculture', *Metadata Semant.*, pp. 481–492, 2007.

[107] D. Tudhope, H. Alani, and C. Jones, 'Augmenting Thesaurus Relationships: Possibilities for Retrieval', *J. Digit. Inf.*, vol. 1, no. 8, Feb. 2001.

[108] S. Schulz, D. Schober, I. Tudose, and H. Stenzhorn, 'The Pitfalls of Thesaurus Ontologization–the Case of the NCI Thesaurus', in *AMIA Annual Symposium Proceedings*, 2010, vol. 2010, p. 727.

[109] B. J. Wielinga, A. T. Schreiber, J. Wielemaker, and J. A. C. Sandberg, 'From thesaurus to ontology', in *Proceedings of the 1st international conference on Knowledge capture*, Victoria, British Columbia, Canada, 2001, pp. 194–201.

[110] M. Hepp and J. de Bruijn, 'GenTax: A generic methodology for deriving OWL and RDF-S ontologies from hierarchical classifications, thesauri, and inconsistent taxonomies', in *The Semantic Web: Research and Applications*, Innsbruck, Austria, 2007, pp. 129–144.

[111] N. F. Noy and D. L. McGuinness, 'Ontology Development 101: A Guide to Creating Your First Ontology', Stanford University, Stanford, U.S.A., Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, Mar. 2001.

[112] S. Staab, R. Studer, H. P. Schnurr, and Y. Sure, 'Knowledge processes and ontologies', *Intell. Syst. IEEE*, vol. 16, no. 1, pp. 26–34, 2001.

[113] M. Uschold and M. King, 'Towards a methodology for building ontologies', in *Workshop on basic ontological issues in knowledge sharing*, 1995, vol. 74.

[114] L. Jansen and S. Schulz, 'The Ten Commandments of Ontological Engineering', in *Proceedings of the 3rd Workshop of Ontologies in Biomedicine and Life Sciences (OBML), Berlin, 6.-7.10.2011*, 2011.

[115] N. Guarino and C. Welty, 'A formal ontology of properties', in *Proceedings of EKAW-2000*, Berlin, 2000, vol. LNCS Vol. 1937, pp. 191–230.

[116] N. Guarino and C. Welty, 'Towards a Methodology for Ontology Based Model Engineering', in *Proceedings of International Workshop on Model Engineering IWME2000*, Nice, France, 2000.

[117] N. Guarino and C. Welty, 'Evaluating ontological decisions with OntoClean', *Commun. ACM*, vol. 45, no. 2, p. 65, 2002.

[118] N. Guarino and C. Welty, 'Identity and Subsumption', in *The Semantics of Relationships: An Interdisciplinary Perspective*, R. Green, C. A. Bean, and S. H. Myaeng, Eds. Kluwer Academic Publishers, 2002.

[119] A. García, K. O'Neill, L. J. Garcia, P. Lord, R. Stevens, O. Corcho, and F. Gibson, 'Developing Ontologies within Decentralised Settings', in *Semantic e-Science*, vol. 11, H. Chen, Y. Wang, K.-H. Cheung, R. Sharda, and S. Voß, Eds. Springer US, 2010, pp. 99–139.

[120] M. Fernández-López, A. Gómez-Pérez, and N. Juristo, 'Methontology: from ontological art towards ontological engineering', 1997.

9. Appendixes

Appendix 1: Source of the steps for the re-engineering method

In section 2 we detailed that the steps in our re-engineering method are the results of the practical application of a naive re-engineering method. The steps in the naive re-engineering method stem from (a) a theoretical comparison of thesauri and ontologies [34] and (b) an analysis of general ontology engineering literature. The theoretical comparison revealed the following steps:

a) Distinction of thesaurus concepts
b) Distinction of thesaurus relationships
c) Sub-distinction of whole-part and associative relationships

Our analysis of general ontology engineering literature revealed several steps that are content-focused as well as precise and actionable. These steps are summarized in table 6, which also lists the respective authors and publications.

Table 6. 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 [11], OBO Foundry principle under discussion, Borgida and Brachman [86]
b) Establishment of an is-a hierarchy	Noy and McGuinness [111], Borgida and Brachman [86], Staab et al [112]
c) Alignment to a top-level ontology	Uschold and King [113], Smith and Ceusters [11], OBO Foundry principle under discussion, Jansen and Schulz [114]
d) Application of the OntoClean method	Borgida and Brachman [86], Guarino and Welty [102], [115]–[118],
e) Establishment of a single inheritance hierarchy	Rector [76], Smith and Ceusters [11], 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 [86], 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 [119], Noy and McGuinness [111], Borgida and Brachman [86], Staab et al [112]
h) (Correct) Codification in a formal representation language	Uschold and King [113], Fernández-López et al [120], García et al [119], Borgida and Brachman [86], Staab et al [112], Rector et al [87], Accepted OBO Foundry principle
i) Provision of metadata for all classes and relationships such as textual definitions and labels	Accepted OBO Foundry principle, Jansen and Schulz [114]
j) Adhering to naming convention for the labels	Accepted OBO Foundry principle, Jansen and Schulz [114]
k) Delineation from existing ontologies	Smith and Ceusters [11], Accepted OBO Foundry principle

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

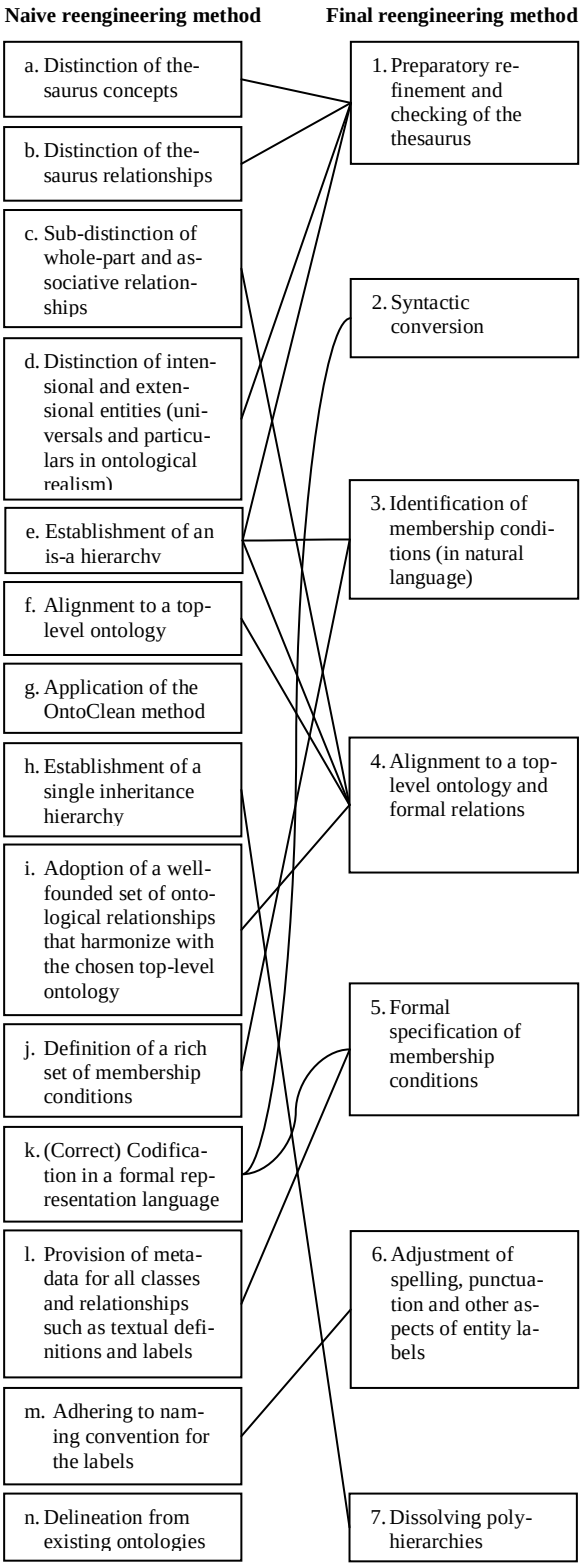

Figure 10. Relation between steps in the naïve reengineering method and the final reengineering method.

Appendix 2: Overview of the steps and subactivities of the reengineering method

The re-engineering method explained in section 2 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. Syntactic conversion
   a. Choice of a formal language
   b. Choice or development of conversion tools*
   c. Conversion of the thesaurus into the formal language
3. Identification of membership conditions (in natural language)
   a. Collection of definitions in natural language**
   b. Extraction of membership conditions
4. Alignment to a top-level ontology and formal relations
   a. Choice of an existing top-level ontology and formal relations
   b. Choice of relevant existing domain-specific ontologies*
   c. Amendment of the thesaurus or the external ontologies*
   d. Alignment of the thesaurus to the top-level ontology
   e. Alignment of the referenced domain-specific ontologies to the top-level ontology*
   f. Alignment of the newly introduced classes to the top-level ontology*
5. Formal specification of membership conditions
   a. Choice of an ontology editor and reasoning algorithm*
   b. Formalizing the class specifications
   c. Adding natural language definitions and comments as class annotations*
   d. 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, its storage format and storage system (steps 1 and 2.b), the availability and quality of existing ontologies (step 4.b/c/e/f), currently existing tools (step 5.a), 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 of the fertilizer concepts in the AGROVOC thesaurus during the case study. The respective step was discussed in section 3.1. 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

Figure 11. Arrays identified for fertilizers and subordinated concepts in the AGROVOC thesaurus and the node labels indicating the arrays

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 3.3. 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 3.4. Table 8 also indicates the plant nutrient levels extracted from the fertilizer regulation by the European Commission [45] and the German fertilizer regulation "Düngemittelverordnung" [47].

We chose to stick to natural language formulations in describing 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 7. Membership conditions of the fertilizer class and its subclasses

Class/ fertilizer type	**Membership conditions**	**Necessary/ sufficient[A]**
Fertilizers	being a compound of collective material entities bearing the disposition to release plant nutrients having a component that has a minimal mass proportion of 1680 ppm plant nutrients as granular part	necessary
fertilizer types listed in table 8, e.g. calcium fertilizer	being a fertilizer having a component that has the minimal mass proportion of a plant nutrient (chemical atom) as granular part as indicated in table 8, e.g. the mass proportion of 143,000 ppm calcium bound in some molecule containing calcium	see table 8
Compound fertilizers	being a fertilizer 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 having a component that has a mass proportion of 1680 ppm plant micronutrients as granular part	necessary

Class/ fertilizer type	Membership conditions	Necessary/ sufficient[A]
Organic fertilizers	being a fertilizer having a component that has a significant mass proportion of a carbon-based molecule as granular part	necessary
Biofertilizers	being a fertilizer 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 + sufficient
Composts	being a fertilizer being the outcome of a decomposition process in which the agent is some living organism and the patient is a dead body	necessary + sufficient
Fish manures	being a fertilizer being the outcome of a crushing or powdering process in which the patient is 'dried fish rest'; *'dried fish rest'* is defined as the outcome of a drying process in which the patient is 'fish rest'; *'fish rest'* is being defined as the dead body of fish or physical parts thereof	necessary + sufficient
Green manures	being a fertilizer being the outcome of a decomposition process in which the agent is a living organisms and the patient is the dead body of a plant or a physical part thereof	necessary
Guano	being a fertilizer 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 not being an organic fertilizer	necessary
Organo-mineral fertilizers	being a fertilizer having some organic fertilizer as a component having some inorganic fertilizer as a component	necessary
Liquid fertilizers	being a fertilizer being a liquid material	necessary + sufficient
Liquid gas fertilizers	being a fertilizer being a gaseous material	necessary + sufficient
Slow release fertilizers	being a fertilizer bearing the disposition to release plant nutrients slowly	necessary + sufficient
fertilizer pesticide combinations	being a compound of collective material entities[B] having a significant mass proportion of fertilizer as component 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 8. Necessary parts of element- or molecule-focused fertilizers in relation to ChEBI

Class/fertilizer type	Granular part (as defined in ChEBI)	ppm	Atom	Necessary/ sufficient[A]
Calcium fertilizers	'calcium molecular entity'	143000	Ca	necessary + sufficient
NPK fertilizers	'phosphorus molecular entity' and 'potassium molecular entity' and 'nitrogen molecular entity'	654 2075 5000	P K N	necessary + sufficient
Nitrogen phosphorus fertilizers	'phosphorus molecular entity' and 'nitrogen molecular entity'	654 5000	K N	necessary + sufficient
Nitrophosphates	'calcium hydrogenphosphate' and 'ammonium nitrate' and 'diammonium hydrogen phosphate'	n/a		necessary
Nitrogen potassium fertilizers	'potassium molecular entity' and 'nitrogen molecular entity'	2075 5000	K N	necessary + sufficient
Phosphorus potassium fertilizers	'phosphorus molecular entity' and 'potassium molecular entity'	654 2075	P K	necessary + sufficient
Magnesium fertilizers	'magnesium molecular entity'	84600	Mg	necessary + sufficient
Phosphate fertilizers	'phosphorus molecular entity'	30956	P	necessary + 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 + sufficient
Sulphur fertilizers	'sulfur molecular entity'	55000	S	necessary + sufficient
Nitrogen fertilizers	'nitrogen molecular entity'	45000	N	necessary + sufficient
ammonium fertilizers	'ammonium compound'	n/a		necessary
nitrate fertilizers	'nitrates'	n/a		necessary

[A] Primitive class with necessary conditions or Defined class with necessary and sufficient conditions

[B] Apatite represents a collective material in the ChEBI ontology so that no reference to the granular is necessary

.

Table 9. 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 + sufficient
plant micronutrient	(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 + 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] Primitive class with necessary conditions or Defined class with necessary and sufficient conditions

Appendix 5: Adjustments of labels (step 6)

The table shown here presents the result of adjustments the class labels in the course of the reengineering of the fertilizer concepts and fertilizer-related concepts in the AGROVOC thesaurus. The step was described in section 3.6.

Table 10. Comparison of class labels to former (preferred) terms in the AGROVOC thesaurus

Preferred term for the concept in the AGROVOC thesaurus	Label for the class 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