Applied Ontology 7 (2012) 401–428
DOI 10.3233/AO-2012-0118
IOS Press

Relationships and relata in ontologies and thesauri: Differences and similarities

Daniel Kless *, Simon Milton and Edmund Kazmierczak
University of Melbourne, Melbourne, Australia

Abstract. Ontologies and thesauri structure concepts and, from the perspective of a practitioner, do not appear to be very different. Nevertheless, experts acknowledge that ontologies are different from thesauri in several respects. In this paper we aim to clarify some of the similarities and differences by systematically comparing the structure of each: the relations and their relata in ontologies and the relationships and their relata in thesauri. In particular, we analyze thesaurus relationships and their relata as they are defined in the latest version of the international (ISO) thesaurus standard against formally well-defined ontological relationships and relata from ontology literature – more specifically ontology literature based in realism. We have found that the relata as well as the relationships in thesauri need to be classified further before any reasonable matching to formal ontological relationships is possible. Isolated hierarchical relationships in thesauri then may correspond to the *is-a* relationship, specific mereological relationships, or fundamental relationships such as the instantiation between universals and individuals in ontologies. Determining how such correspondences apply in domain-specific cases depends on whether the thesaurus relationships contribute to the specifications of necessary and sufficient conditions for their respective relata in the ontology – a function that relationships do not have in thesauri. Our findings make it clear that thesauri require structural and definitional reengineering in order to be reused or treated as ontologies, but that adherence to the international standard for thesauri provides a good base for such reengineering.

Keywords: Ontologies, terminologies, thesauri

1. Introduction

According to a 2007 online summit "Ontology, taxonomy, folksonomy – Understanding the distinctions" large parts of the ontology community describe ontologies on a spectrum of artifacts that ranges from formal upper-level ontologies expressed in first order logic over taxonomies, through other structured vocabularies to folksonomies (Gruninger et al., 2008). This spectrum is frequently discussed in the literature (Guarino et al., 2009; Obrst, 2006; Uschold & Grüninger, 2004). While being easy to display, it is questionable, if it is the formalism and the language that explains the difference between ontologies and structured vocabularies.

Some authors refer to "terminologies" when discussing structured vocabularies from an ontological perspective. One criticism in such discussions is the uncertainty that inheres in the conceptual framework of terminologies (Smith, 2004; Smith et al., 2005). This criticism does not aim at the formalism, but on the content of terminologies. While the criticism focuses on shortcomings of the conceptual framework in general terminologies, it sheds little light on the true similarities and differences.

With this paper we aim to clarify the differences and similarities between thesauri and ontologies and lay the basis to guide the choice between the two. We do so by restricting our attention to *thesauri* as a

*Corresponding author: Daniel Kless, University of Melbourne, Melbourne, Australia. E-mail: klessd@unimelb.edu.au.

specific type of structured vocabulary and then giving a detailed comparison of thesauri to ontologies. In particular, we aim to analyze the differences and commonalities between thesauri and ontologies in terms of (a) the *relationships* that each provides and (b) what these relationships relate – their *relata*. These entities are of fundamental importance in ontologies as well as thesauri and address what is generally first perceived when looking at domain-specific examples of these artifacts.

The comparison we pursue here is a theoretical one, based on a specific set of ontology literature and a specific thesaurus standard. We will neither elaborate on differences between thesaurus standards nor differences between formal specifications of ontological relationships. We will also not pay attention to how these relationships have been adopted and/or adapted in practice. The theoretical focus has the advantage that the results have a general character, unspecific to a particular implementation. We acknowledge that there can be enormous differences between the recommendations in thesaurus standards and practice. In that sense we deal with thesauri as they *should* be constructed.

Section 2 provides some background on ontologies, thesauri, terminologies and structured vocabularies in general. The paper is then organized into three main parts. The first one (Section 3) pays attention to what relationships in thesauri and ontologies relate respectively – their relata. The second part (Section 4) analyzes thesaurus relationships from an ontological perspective with a focus on the formal specification of the relationships. Section 5 concludes the paper. The specific method used for each analysis is introduced in the relevant sections.

2. Background: Ontologies, thesauri and their comparison

2.1. Ontologies

Today there can be distinguished at least two views on ontologies, which Poli and Obrst (2010) named (1) ontology as categorial analysis and (2) ontology as technology. Ontology as *categorial analysis*, generally referred to as formal ontology, is concerned with the general metaphysical analysis of reality in the sense of defining categories of existence such as *thing*, *process*, *matter*, *form*, *whole* and *part*. In this context formal systems, such as first-order logic, are often used to formalize theories of reality in which reasoning about the ontology can take place. Ontology *as technology* is generally seen as the result of ontological engineering and pursuing the purpose of knowledge representation. It has a strong focus on computational aspects and is often associated with formal languages such as the Web Ontology Language OWL (OWL2-Overview, 2009).

Poli and Obrst highlight that both viewpoints, the one of categorial analysis and the technology perspective, can benefit from each other. Other authors (e.g., Bittner & Donnelly, 2007) argue that the respective strengths of the applied formalism are useful to exploit. Nevertheless, there is still little clarity on how categorial analysis, as a process, is appropriately applied using the technology developed in ontology engineering and knowledge representation. In particular, there is little clarity if, and how, entities such as classes, individuals, resources, universals and particulars used in the respective approaches to ontology match or subsume each other. The confusion is even greater since resources such as the Gene Ontology are in an ongoing process of improvement through categorial analysis (Smith & Ceusters, 2010), are referred to as "controlled vocabulary"[1] and published using OWL or other knowledge representation languages[2] that are generally associated with ontologies.

[1] http://www.geneontology.org/ (status of the start page: 7.2.2012).

[2] http://www.geneontology.org/GO.downloads.ontology.shtml.

We do not aim to address this specific uncertainty, but will approach ontologies from a philosophical orientation where categorial analysis is central and logic is used to formalize the categories and relationships between the categories. This role of logic is described as a theory of science and traced back to Husserl and his predecessor Bolzano (Smith, 1989). The respective philosophically oriented literature allows us to provide a proper grounding for what is represented in ontologies – an aspect that is often not paid much attention to in technology-focused approaches to ontology.

2.2. Thesauri

In contrast to ontologies, where only representation languages have been defined at this point of time, the conceptual framework of thesauri is comprehensively defined in various standards, most notably part one of the international thesaurus standard (ISO 25964-1:2011). This standard is the first update since its predecessors (ISO 2788:1974; ISO 2788:1986) have been published 25 years earlier, but the way the current standard deals with relationships has not changed significantly. The existence of an international thesaurus standard does not imply that it provides a unique perception of thesauri. Although the standards are visibly coordinated, various national standardization committees did not simply adopt and translate the ISO standard, but modified the rules in ways that often go beyond cultural or language-specific modifications, e.g., ANSI/NISO Z39.19-2005 in the USA or DIN 1463-1:1987 in Germany. In this paper we will pay particular attention to the ISO standard since the definitions and rules are certainly the furthest consent that exists between experts in the field.

In the ISO standard a thesaurus is defined as a "controlled and structured vocabulary in which concepts are represented by terms, organized so that relationships between concepts are made explicit, and preferred terms are accompanied by lead-in entries for synonyms or quasi-synonyms"[3] (ISO 25964-1:2011, Section 2.62).

Many information retrieval thesauri are limited to a certain scope (e.g., medieval literature or biology in general) that may not only limit the breadth of terms covered by a thesaurus, but also its level of detail/specificity, and the meaning which is typically given to the terms. For example, the notion of water may be defined differently in chemistry (a molecule) and agriculture (a resource). In this and other aspects thesauri as they are defined in standards differ from thesauri as general-purpose word-finding dictionaries (such as Roget's Thesaurus). Information retrieval thesauri have a considerable history reaching back to the 1960s (Aitchison & Clarke, 2004) – far before computers became widespread as knowledge processing instruments and their underlying logics was considered a potentially useful influence for formalizing knowledge.

2.3. Terminologies, vocabularies and knowledge organization systems

While our paper is focused on comparing thesauri and ontologies, part of our comparison in Section 3 requires the consideration of terminologies. Similar to the term "ontology", the term "terminology" is used in a variety of ways in common practice. Often it refers to what linguists describe as a concept system, a "set of concepts structured according to the relations among them" (ISO 1087-1:2000, Section 3.2.11). Such structured sets of concepts are better known as types of **structured vocabularies**, an "organized set of terms, headings or codes representing concepts and their inter-relationships, which can

[3] A *controlled* vocabulary is defined as a "prescribed list of terms, headings or codes, each representing a concept" (ISO 25964-1:2011, Section 2.12), a *structured* vocabulary a "organized set of terms, headings or codes representing concepts and their inter-relationships, which can be used to support information retrieval" (ibid., Section 2.56).

be used to support information retrieval" (ISO 25964-1:2011, Section 2.56) or **knowledge organization systems**, schemes for "organizing information and promoting knowledge management" (Hodge, 2009; see also Chowdhury & Chowdhury, 2007; SKOS-Reference, 2009) and include thesauri, classification schemes and taxonomies. In this paper we prefer using the term "structured vocabulary" for this particular understanding of the term "terminology". The meaning of "structured vocabulary" overlaps with "**controlled vocabulary**" that refers to a "prescribed list of terms, headings or codes, each representing a concept" (ISO 25964-1:2011, Section 2.12). Controlled vocabularies do not have to be structured. Often simply the term "**vocabulary**" is used to summarize structured and unstructured vocabularies.

In this paper we will refer to "**terminology**" as a type of vocabulary, such as glossaries and termbases, "used for unambiguous communication in natural, human language" (ISO 704:2009, Section 0.1). This understanding is in accordance to its definition in the international terminology standard (ISO 1087-1:2000, Section 3.5.1) as a "set of designations belonging to one special language" (the terms used in a subject field). Terminologies in this sense are highly standardized, most prominently in standards concerned with terminology work published by the International Standardization Organization (ISO 1087-1:2000; ISO 704:2009; ISO 860:2007). The way the terminology standard (ISO 704:2009, Section 5.5) suggests applying relationships has parallels to the thesaurus standard (ISO 25964-1:2011, Section 10). When comparing ontology and thesaurus relationships in Section 4 we will not analyze differences and similarities of relationship definitions in the thesaurus and terminology standards, but only focus on relationships as they are defined in the thesaurus standard (ISO 25964-1:2011, 2011).

2.4. Existing comparisons of thesauri and ontologies

Existing comparisons of ontologies, thesauri and other types of structured vocabularies are very limited in terms of the insights they provide. The stream of ontology literature that compares ontologies with other types of structured vocabularies is based on language formalism and expressivity (Gruninger et al., 2008; Guarino et al., 2009; Obrst, 2006; Uschold & Grüninger, 2004) and does not explain, what effect the formalism has or should have on the ontology content or its quality, if there is such effect at all.

Another stream of literature that relates structured vocabularies to ontologies can be roughly summarized under the label "knowledge organization" (BS 8723-3:2007, 2007; Chowdhury & Chowdhury, 2007; Gilchrist, 2003; van Rees, 2003). The publications are mainly focused on introducing the variety of "classical" knowledge organization systems (e.g., classification schemes, thesauri, authority files) and consider ontologies as a new type of knowledge organization system. These comparisons are generally too superficial to give a deep understanding of the similarities and differences between thesauri and ontologies and often present ontology as technology. None of the publications above addresses structure or content of ontologies and how they differ, or are common to, other knowledge organization systems.

Further, there have been attempts to reengineer thesauri into ontologies, which require some stance as to what the differences between the two actually are. An example is the AGROVOC thesaurus,[4] the reengineering of which[5] was driven by the stance that ontologies would allow a finer distinction of relationships than the default relationships provided by thesauri and that ontologies can be more precise about the meaning of terms and concepts (Soergel et al., 2004). While these and other aspects can be differences between thesauri and ontologies, they are not systematically addressed by Soergel et al. and just weakly relate to ontology as categorial analysis.

[4]http://aims.fao.org/website/AGROVOC-Thesaurus/sub.

[5]Results accessible under http://code.google.com/p/agrovoc-cs-workbench/.

Categorial analysis has played a bigger role in recent development of biomedical vocabularies such as the Gene Ontology[6] or the NCI Thesaurus[7] (Munn & Smith, 2009; Smith & Ceusters, 2010). Nevertheless, there has been critique that the methodology behind it is still not clear and rather made untraceable by attributing it to the philosophical stance of ontological realism (Merrill, 2010). A general expectation of using formalism is the ability to reason about the ontological knowledge. There have also been warnings that typical expectations of ontologies such as logic-based reasoning would not necessarily lead to useful results in the current state of the NCI Thesaurus (Ceusters et al., 2005; Schulz et al., 2010).

Overall, the discussions show that there are either paradigmatic debates on the differences between ontologies and thesauri or comparisons remain shallow. We particularly find that there is a lack of precise comparisons. Relationships and the things that they relate require a more systematic analysis and deeper understanding than accounted for in the reengineering of the AGROVOC thesaurus for instance.

The work of Guarino and Welty (2009) on the *OntoClean* method is also relevant to our discussion despite not being a direct comparison of thesauri and ontologies. The method argues for correctness of ontological hierarchical subsumption based on some rules that relate to the four philosophical notions (meta-properties) of essentiality/rigidity, identity, unity and dependency. This method has been applied to ontologically improve Wordnet (Oltramari et al., 2002). In contrast to OntoClean, the approach we follow in this paper is less focused on the hierarchical subsumption in ontologies alone, but rather on the full range of comparable thesaurus and ontology relations based on the ontological classification of their relata.

3. Comparison of ontology and thesaurus relata

Our aim in this section is to identify the relata between which relationships exist in thesauri and ontologies, respectively. This is an important first step towards establishing correspondences between relationships in ontologies and relationships in thesauri. In particular we will pay attention to whether relata are intensionally defined. Intensionality plays an important role in logic, particularly with respect to the transitivity of relationships, which will be subject of Section 4.

The first subsection will introduce the comparison approach. Sections 3.2 and 3.3 detail the respective relata of thesauri and ontologies. Section 3.4 lays out the comparison results, Section 3.5 relativizes them from a pragmatic perspective and Section 3.6 finally discusses the results.

3.1. Comparison approach

We compare thesauri and ontologies on a theoretical basis by investigating the relationships provided by each. This necessitates a comparison of the relata within relationships, which in turn requires us to rely on the definitions in the literature. On the side of the ontology literature, we will ground our understanding of ontologies on the literature regarding the philosophical stance of realism. Ontological realism is generally opposed to conceptualism, nominalism and resemblance theories by accepting the existence of universals as opposed to individuals (Woozley, 1967, 2004), both of which will be further discussed in the next subsection. Our choice of realism is because this stance has found wide acceptance

[6]http://www.geneontology.org/.
[7]http://ncit.nci.nih.gov/.

in the field of formal ontology. As this paper will argue, it is useful for comparing thesauri and ontologies since terminologies make a distinction that is at least similar to the one between universals and individuals found in realism ontologies. The identified ontology relata will be detailed in Section 3.2.

In case of thesauri we will build upon the definitions made in part one of the international thesaurus standard ISO 25964-1:2011. The thesaurus relata will be discussed in Section 3.3.

The comparison of the relata we approached by establishing mappings between the relata of ontologies and thesauri. As in many mapping attempts, one must expect exact correspondences, partial correspondences or no correspondences between entities. In case of partial correspondences, it is possible that an entity on one side corresponds to several entities on the other side (one-to-many correspondences).

The correspondences that we establish only express that there are *some* entity instances on one side that correspond to some entity instances on the other side. This implies that one-to-one correspondences may not be exact, but only partial matches and one-to-many correspondences may be incomplete. This information is sufficient for us since we are primarily interested in whether we can compare relationships of thesauri and ontologies at all and whether we need to be aware of important known multiple correspondences.

It is difficult to map thesaurus and ontology relata directly based on their definitions. For this reason we made critical use of terminology standards. The entities defined in these terminology standards can be used as a mediator between thesaurus and ontology entities. Our use of terminology standards will be detailed further in Section 3.4.

3.2. Ontology relata and how they convey meaning

For addressing the relata that can be found in ontologies, we will build upon a reference typology for realism ontologies presented by Kuśnierczyk (2006). The goal of Kuśnierczyk's typology is to refer to the same things in ontologies consistently. It is based on earlier work of Bittner et al. (2004) and served there to explain fundamental ontological relationships, which we will introduce in Section 4.3. Smith et al. (2006) provide another example of a definitional framework which is similar to the one proposed in this paper.

All the things that ontologies deal with are called **entities**. The fundamental types of entities in ontologies that follow the doctrine of realism are *particulars* and *universals*. Seibt (2010) summarized the common reference to **particulars** (also known as **individuals**) as de facto references to Aristotelian substantial particulars that are concrete, individual, logical subjects, determinate, enduring and unique in space at any time (e.g., my car, Charlie Chaplin). Today's perception of particulars also includes entities that fulfill just some of these criteria such as numbers, processes, or social systems (e.g., an economy). A useful minimal criterion for particulars is that they cannot be exemplified or repeated (MacLeod & Rubenstein, 2005).

According to Kuśnierczyk (2006), particulars should be opposed to universals and collections. **Collections** are multiple particulars (e.g., 'the current parliament members' as collection of specific persons). The members of some collections are a number of specific exemplars of one single *common* universal. We will refer to such collections as **pure collections**. There exist also other kinds of collections, the members of which are not exemplifications of a single universal, e.g., the sum of all the specific things that we refer to as vegetables – zucchinis, eggplants, potatoes and so forth – or all the specific things that are edible but not fruits (incl. vegetables, meat and so forth). We will call such assemblies **mixed collections**. Mixed collections often appear to be socially defined – through customs, habits, laws, regulations, common practice and so forth – and opposed to things whose delimitation is built on scientific grounds. A notion that appears similar to mixed collections is what Wittgenstein described as "family resemblance". The concept "game" is used as frequently as the 'vegetable' example (Stock, 2010). Also Armstrong (1978, Chapter 13, Section IV) refers to a "family" when distinguishing it from a universal and other things.

Universals, in contrast to collections and particulars, *can* be exemplified. Collections exist at one point of time only and can lose members while universals are timeless and may have changing members (particulars) or even no members at any given point of time. Two universals may have the same extension (the same members) at a certain point of time, but not any two collections.

Universals *can* be specified *intensionally* (Guarino et al., 2009). Specifically, formal ontologies specify a universal by properties and relationships to allow making membership decisions for particulars using the means of logic. To achieve useful reasoning results that correspond to reality, ontologies specify universals only by properties that are true for all particulars being member (instance) of a universal *under all circumstances and in the entirety of their existence*. Thus, the properties express *necessary conditions* for particulars to be member of a universal (e.g., the ability to think attributed to humans).

Notably, an ontology must *not* state any *accidental* properties of universals that its members may or may not have. For example, 'being able to walk' or 'having a leg' are accidental properties and not necessary conditions of 'humans' since these properties would exclude infants as well as people that had severe accidents or have disfiguring birth defects. All of them are still humans. An ontology that modeled such properties for 'humans' and subsumed 'human infants' and 'disabled people' under 'humans' would lead to reasoning results that do not match reality.

Ideally, a universal is not only specified by necessary properties, but defined through *necessary and sufficient* membership conditions. We will refer to such specifications as *intrinsic definitions*. The properties that form a *necessary and sufficient* membership condition are common to all the particulars represented by the universal and cannot be found in the extension of any other universal. For example, a universal 'triangle' is sufficiently described by the property of being a shape and having three sides that touch each other in exactly one point each. This combination of conditions applies to nothing but triangles. It is an intrinsic definition.

Necessary conditions as well as necessary and sufficient conditions are truth-conditions that ontologies use to establish meaning. These truth-conditions can be efficiently utilized by reasoning algorithms.

3.3. Thesaurus relata and how they convey meaning

It is a general perception that thesauri define relationships between terms. While this is true for past international thesaurus standards, contemporary thesaurus construction separates terms from concepts and most relationships are established between concepts (ISO 25964-1:2011). The thesaurus standard defines a **concept** as a "unit of thought" (ISO 25964-1:2011, Section 2.11); a **term** as a "word or phrase used to label a concept" (Section 2.61). The role of terms is barely more than labeling a concept.

The meaning of thesaurus concepts is conveyed to *human* users by (a) the terms assigned to the concepts combined with the "most common" understanding of the term's meaning, particularly the meaning of the preferred term, sometimes also considering the context of the domain for which the thesaurus is designed and in which users have a particular preoccupation of term meanings (Sections 6.2.1 and 6.2.2), (b) the thesaurus hierarchy (Section 4.2) as a kind of "context" for a thesaurus concept, (c) associated concepts (Section 4.2), some of which indicate what the concept in question is not meant to include – basically a list of "frequently confused concepts", (d) a qualifier, (e) a scope note, (f) a precise definition. Qualifiers are terms added in brackets to the main term, e.g., bank (*financial institution*) or bank (*geography*). They are used to disambiguate the meaning of homographs – terms with identical spelling, but different meanings (Section 6.2.2). Scope notes "should be used to clarify the boundaries of a concept, especially when the meaning of the preferred term in ordinary discourse can be interpreted too broadly or too narrowly, or to distinguish between preferred terms that have overlapping meanings in natural language" (Section 5.2). Definitions are not considered necessary in thesauri (Section 6.2.3).

The thesaurus standard explicitly states that concepts can be "deliberately restricted in scope to selected meanings" and that scope notes "should state the chosen meaning, and it may also indicate other meanings that are recognized in natural language but which have been deliberately excluded for indexing purposes" (Section 4.2). This shows that the meaning, which a thesaurus concept is supposed to convey, may deviate from the "dominating" meaning of the term or terms representing the concept. An example is where the terms assigned to a concept are not true synonyms, but only quasi-synonyms "whose meanings are generally regarded as different in ordinary usage but which may be treated as labels for the same concept", e.g., 'diseases' and 'disorders' (ISO 25964-1:2011, Section 2.47). Quasi-synonyms "frequently represent points on a continuum, and can even be antonyms (opposites)", e.g., 'wetness' and 'dryness' (Section 8.3). Another case of incongruence between term and concept meaning is where the complexity of a thesaurus is reduced by deliberately not distinguishing concepts with more restricted, or narrower, meanings, e.g., the terms 'basalt', 'granite' and 'slate' may all be assigned to an existing concept that is already labeled 'rock' (Section 8.4).

3.4. Results of the comparison of thesaurus and ontology relata

The last two sections left us with a situation where we have universals, individuals and collections on the ontology side, and concepts on the thesaurus side. The definitions alone do not give us any justification for relating entities to each other on any side, not to speak about relationships that can be established between them. To overcome this situation, it is helpful to further analyze these entities and concepts. For this purpose we will use the definitions of a particularly useful terminology standard (ISO 1087-1:2000) to act as an intermediate. The result of this refinement is summarized in Fig. 1.

In Section 2.3 we emphasized that the relatedness of the ISO thesaurus standard to ISO terminology standards is limited and that definitions have been provided independent from each other. Nevertheless, it is well justified to assume a strong correspondence between concepts in thesauri and concepts in terminologies taking into consideration that the terminology standard defines a concept as a "unit of knowledge created by a unique combination of characteristics" (ISO 1087-1:2000, Section 3.2.1). A characteristic is defined as an "abstraction of a property of an object or of a set of objects" (Section 3.2.4).

For the moment we are more interested in the distinction that the terminology standard makes between general and individual concepts. Individual concepts are defined as "concepts which correspond to only one object" (Section 3.2.2); general concepts "correspond to two or more objects which form a group by reason of common properties" (Section 3.2.3).[8] The thesaurus standard does not, in general, distinguish between general and individual concepts. Just on one occasion, for defining the instance relationship (see Section 4.2), a distinction is made between a general concept (called "class") and its instances (ISO, 2011, Section 10.2.4.1), but neither of them is defined in the thesaurus standard. Anyway, the fact that we build upon is that concepts in thesauri can be either individual concepts *or* general concepts in terminologies.

[8]The term "correspond" in the definitions matches our use of the term "represent" as explained below.

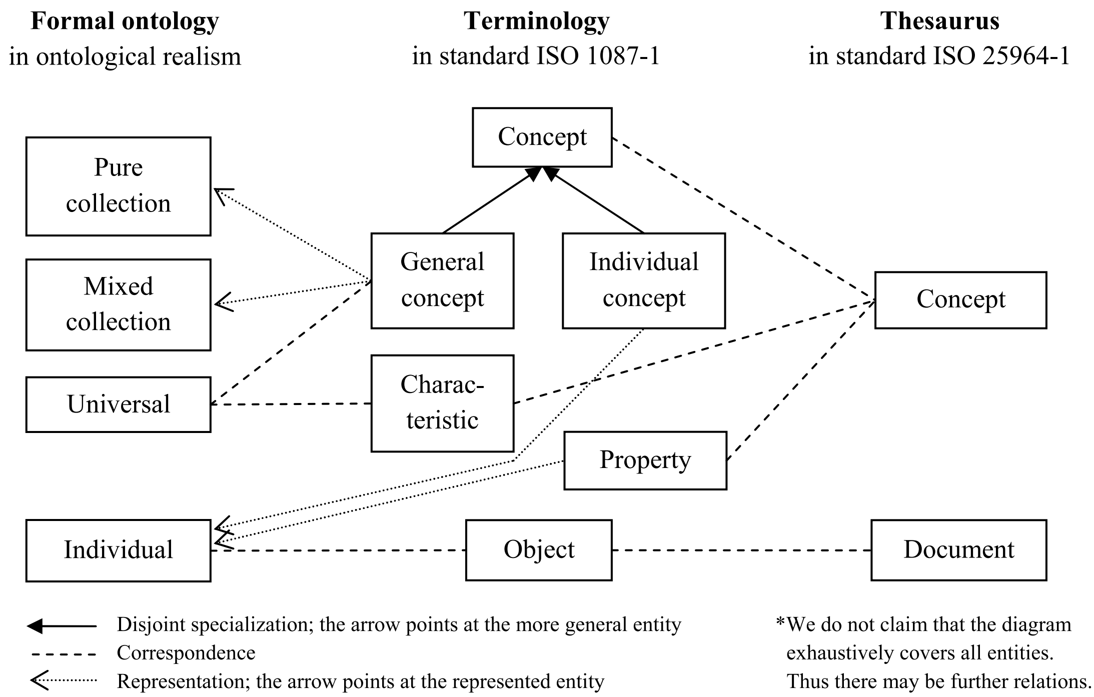

Fig. 1. Identified correspondences between entities in ontologies, thesauri and terminologies.

Properties of objects and thus characteristics of concepts are not specifically the subject of discussion in the thesaurus standard. Just at one point, where a specific type of associative relationship is described (ISO 25964-1:2011, Section 10.3.3, para. e), properties are mentioned, again without defining what properties are. We note that thesaurus concepts may not only correspond to individual or general concepts in terminology standards, but also to their properties or characteristics (e.g., ‘toxicity’ or ‘magnetism’).

An “object” as this term is used in the definitions of general and individual concepts, refers to “anything perceivable or conceivable”, which the corresponding footnote says “may be material (e.g., an engine, a sheet of paper, a diamond), immaterial (e.g., conversion ratio, a project plan) or imagined (e.g., a unicorn)” (ISO 1087-1:2000, Section 3.1.1). The relationship between objects and individuals or general concepts has been described as a “correspondence”, which we interpret as the extension of what the concepts are supposed to represent.

An object in the context of a terminology has at least some correspondence to what is referred to as “document” in the thesaurus standard, which is defined as “any resource that can be classified or indexed in order that the data or information in it can be retrieved” including “written and printed materials [. . .], but also to non-printed media such as machine-readable and digitized records, Internet and intranet resources, films, sound recordings, people and organizations as knowledge resources, buildings, sites, monuments, three-dimensional objects or realia; and to collections of such items or parts of such items” (ISO 25964-1:2011, Section 2.15). Nevertheless, thesaurus concepts can express various things about documents – generally their subject – and thus often do not represent these documents as entities. We will not further elaborate on this correspondence here, since documents are not subjects of relations in thesauri.

We can now turn our attention to building correspondences between the entities defined in terminologies and the entities found in ontologies. Sometimes we cannot only use correspondences to express the relatedness of these entities, but need to make use of a different relation that expresses the *representation of* entities in ontologies through entities in terminologies (see Fig. 1). Note that pure collections and mixed collections are displayed closer to universals in figure 1 only for reasons of readability, while in fact they are more comparable to individuals for being specific instances and not abstractions of them like universals.

Some individual concepts in terminologies are *representations* of individuals in ontologies. General concepts can *correspond* to universals in ontologies, since both universals (see Section 3.2) and concepts in terminologies (ISO 704:2009, Sections 5.4.3 and 6.2) are concerned with providing *intensional* definitions. Nevertheless, the understanding of intensionality in the terminology standard is looser than the one for ontologies. The terminology standard clarifies that terminologies only distinguish shared and delimiting characteristics (ISO 704:2009, Section 5.4.4), but do not pay attention to necessary, sufficient, and essential characteristics (ibid., Section 5.4.5).

The difference becomes obvious in an example given in the terminology standard for the harmonization of concepts and terms (ISO 860:2007). The example was concerned with defining a concept 'domestic wastewater'. After analyzing various dictionary and encyclopedia definitions of this term, it is found that (ibid., p. 7):

> the characteristics used to define these concepts are not exactly the same (e.g., some state that domestic wastewater comes only from households while others also include the wastewater from business buildings, commercial buildings, institutions and the like), and therefore the intension and the extension of the concepts do not coincide, the concepts are closely related and there is an important area of overlap so that by adjusting the essential and delimiting characteristics, the concept can be harmonized.

This situation should lead to the definition of different universals in an ontology. The terminology standard, instead, suggests the following "harmonized" definition for 'domestic wastewater' (ibid., p. 8):

> waste-water from household, business buildings or institutions.

To qualify as an intrinsic definition of a universal (as defined in Section 3.2), the question would have to be answered, what being sourced from households, business buildings or institutions have in common that, e.g., being sourced from rainwater drainage systems does not have. The debate would soon lead to the question, whether water from households has essential properties that differ from the essential properties of water from business buildings or from institutions.

We do not want to deepen the discussion of this specific example, which may be argued either way. We simply take it as a case in which the represented collection of things is not exemplifying a common universal. Instead, it is what we described as a mixed collection. We also assume that terminologies do not only contain concepts that are *representations* of mixed collections, but also are *representations* of pure collections, although this might be rather rare.

Since the thesaurus standard does not deal with properties and characteristics at all, it also makes no commitment as to whether its concepts are intensionally or extensionally defined. There is one (optional) feature in thesauri, which can at least support making some of the intensional characteristics of concepts transparent: array labels (often called facets). Array labels, when applied within the hierarchy of a thesaurus, indicate the "characteristic of division" by which a group of sibling concepts are distinguished (ISO 25964-1:2011, Section 2.3.8). The following example, illustrates several array labels (indicated in italic font) using the concept of 'water':

- Water *by location*: ground water, surface water.

- Water *by pollution level*: sewage, recycled water, freshwater.
- Water *by origin*: industrial water, household water, rain water, tap water.
- Water *by salinity*: salt water, freshwater.
- Water *by purpose*: irrigation water, drinking water.

The presence of array labels does not yet reveal, what the intension of the concept is, but only that there are delimiting properties that all have some commonality. This commonality may be described as a common superordinate property and is often called an "attribute" in conceptual data modeling.

At some point we noted that thesaurus concepts may not only *correspond* to terminology concepts, but also to properties or characteristics in terminologies. Characteristics (e.g., being red) can *correspond* to universals in ontologies. Properties in terminologies (e.g., the redness of a specific carpet) can *represent* individuals in ontologies. Also individual concepts in terminologies can *represent* individuals in ontologies. Finally, objects as they are defined in the terminology standard all *correspond* to individuals in ontologies. There could be further discussion whether individuals need to be "perceivable or conceivable", but we consider such discussions overly meticulous and one may well argue that the precise definition of individuals is not without debate in philosophy (Seibt, 2010).

Following the logic of this section (as displayed in Fig. 1), but also on reasonable grounds one can expect that some thesaurus concepts *correspond* to intensionally defined universals, some *represent* individuals, some *represent* pure collections and some *represent* mixed collections.

3.5. Pragmatic aspects of matching specific instances of thesaurus and ontology relata

In our comparison we were mainly concerned with establishing correspondences between relata in thesauri and ontologies from a meta-perspective. If a correspondence between a thesaurus concept and a universal in an ontology is a valid mapping in a specific instance depends on whether both convey the same meaning, i.e. whether the instances of a universal are *exactly* the same as the objects that a thesaurus concept (and terminology concept) represents.

Assessing the validity of a specific mapping requires mediating between the different principles of expressing meaning in thesauri and ontologies as they were laid out in Sections 3.2 and 3.3. Such assessment also depends on the quality of the thesaurus and ontology. It is a fundamental problem that many thesauri do not contain definitions, since ISO 25964 neither considers them necessary nor offers any rules for definitions. The lack of definitions makes thesaurus concepts prone to different interpretations or to confuse intensions. For example, a concept 'water' in a thesaurus may have – even in a specific domain – different ontological interpretations, e.g., as a chemical substance with certain elements, or as a liquid made out of that substance (i.e., 'water' in a certain temperature range, excluding 'ice', 'water steam' or 'plasma'). This kind of blurring of the intension of concepts may often be unconsciously made in daily thesaurus construction and maintenance work – or even intentionally to subsume quasi-synonyms (see Section 3.3) under one concept.

Ontologies convey meaning only to the extent to which membership conditions have been defined through properties – ideally stating necessary and sufficient conditions. Interestingly, as far as the ISO standards are concerned, terminologies are ahead of thesauri in that they are expected to model properties. Terminologies are thus closer to ontologies than thesauri with respect to modeling properties. Nevertheless, we are not familiar with any ontology modeling essential or at least necessary conditions using properties (attributes and relations) for the majority of its universals – at least not in a size of thousands or ten thousand distinguished entities as it would be typical for domain thesauri. This lack of specifying conditions appears to be a very weak side of many currently existing ontologies. It is thus often left to means similar to those in thesauri to convey the meaning of universals in an ontology, e.g., natural language definitions, examples or the terms labeling the universals.

3.6. Discussion of comparison results

The comparison of thesaurus and ontology relata revealed that thesauri structure only concepts, each of which may correspond to (a) an intensionally defined universal, (b) a representation of a single individual, (c) a representation of a pure collection or (d) a representation of a mixed collection in an ontology. Whether a concept is an intensionally defined universal makes an important difference in terms of exploiting transitivity of relationships, the subject of next section. If a concept does not denote objects that have unique characteristics, there are also no characteristics that can be inherited by subordinate concepts (hyponyms). The absence of intensional characteristics makes a knowledge model also uneconomical by expressing characteristics about collections of things at a specific state of affairs only – not in a general way.

At this stage we must expect that mapping thesaurus relationships to ontological relationships will require the distinction of concepts into universals or other fundamental entities. This may be very difficult to do posterior, particularly given the fact that concept definitions are not considered a necessary part of a thesaurus and do not have to make explicit the characteristics of concepts. Further, mapping specific instances of concepts and universals requires mediating between two fundamentally different ways of expressing meaning in thesauri and ontologies, i.e., two different ways of conveying the intension of the relata have to be bridged.

4. Comparison of ontology and thesaurus relationships

The first part of the paper (Section 3) paid attention to the relata of relationships in ontologies and thesauri. The goal of this section is to compare structural relationships in ontologies and thesauri, by focusing on two logical aspects of relationships: (a) transitivity of the relationships and (b) the category membership of the relata in relationships.

The following section explains our comparison approach. Sections 4.2 and 4.3 introduce the relationships of the thesauri and ontologies, respectively. The result of comparing these relationships is presented in Section 4.4. Section 4.5 amends this comparison by pragmatic aspects that distinguish the way thesaurus and ontology relations are applied. Section 4.6 discusses the comparison results.

4.1. Comparison approach

In comparing thesaurus and ontology relationships, we selected a limited range of relationships on both sides. In case of thesauri, we can again rely on relationships as they are defined in the international standard ISO 25964-1:2011 including the constraints surrounding their correct application spread across the standard. Although the standard permits the definition of custom relationships (ibid., Section 10.4) and thus leaves an open end for the number of relationships, it lists a set of "core" relationships that are characteristic for thesauri as such. We will focus solely on these relationships when we detail thesaurus relationships in Section 4.2.

Until now, there is no standard that would define ontological relationships comparable to the one for thesauri. Further, at this point of time formal ontology is understood as the activity of defining relationships using axioms rather than a certain range of relationships. Nevertheless, the range of relationships that is typically discussed in connection with ontologies is, apparently, roughly comparable to the core relationships in thesauri as well as the relationships used in other disciplines such as conceptual data modeling. Often these structural relationships are called the "hierarchy" or the "taxonomy" of an ontology.

As much as we focused on a specific set of thesaurus relationships in a single standard – not comparing them with other definitions – we also focus on a fixed set of ontological relations. As a counterpart for the thesaurus standard, we will make use of a limited number of papers that have already summarized ontological relationships. The papers were chosen so that their combination (a) provides a sufficient range of relationships for comparison to thesauri; (b) indicates the relata types and transitivity for these relationships and (c) types the relata through categories that harmonize with each other by adhering to a common philosophical basis – in our case ontological realism.

The first range of ontological relations that we will refer to are the ones by Lowe (2005) combined with, and extended by, some definitions of Bittner et al. (2004). Only the latter authors specified relationships formally using first-order-logic. Lowe's categories and relationships do well in explaining and detailing the relationship between universals and particulars, two of the four fundamental ontological entities that we distinguished in Section 3.2, and which we pay attention to.[9]

A second range of ontological relationships that we refer to focuses on part-of relationships and is based on a paper by Keet and Artale (2008). They have analyzed part-of relationships based on a breadth of papers and defined the relationships formally using first-order-logic. Further, they used the categories of DOLCE to distinguish relationships and specify their relata. DOLCE is a top-level ontology that has a cognitive bias in the sense that it is human language and common sense that underlie its categories (Gangemi et al., 2002, Section 2).

We have deliberately restricted ourselves to comparing relationships in thesauri and ontologies with respect to transitivity and the constraints on the categories from which the relata must come, despite the fact that there are many other possibilities for specifying the logical properties of relationships axiomatically in formal ontology (Bittner, 2009; Bittner & Donnelly, 2007). The focus on transitivity and the typing of relata in relationships does not only scope the paper, but also allows us to examine the construction of hierarchies and claims about reasoning.

Transitivity plays an important role in logic-based reasoning, particularly in combination with the intensional definition of the relata we saw earlier. Transitivity holds where, if a is related to b, and b is related to c, then a is also related to c. If the latter is not true, then the relationship is not transitive. We focus on transitivity because it is a major guiding principle in building good hierarchical structures in ontologies. Stock (2010) claims that transitivity plays an important role for search expansion – one of the information retrieval purposes of thesauri and other knowledge organization systems. Our paper challenges such a view, which is justified, since transitivity has never played an explicit role in the definition of relationships in thesaurus standards.

Constraining the **category membership** of the relata in a relationship refers to the kind of entities between which the relationship may hold (e.g., to say that the relationship 'orbits around' can only apply between two massive physical objects). Category membership of the relata in a relationship is an important guiding principle in applying relationship correctly.

In formal ontology the term "category" is generally used to set apart the most fundamental entities in existence. Categories do not only support constructing hierarchies, but clearly constrain and therefore characterize relationships. Categories of this sort are intended to be mutually exclusive and exhaustively

[9] We will not further elaborate on relationships to categories. Interested readers we refer to Bittner et al. (2004) in this regards.

cover the entities found in reality. However, categories may be defined more specifically for a domain that does not encompass all entities. Categories in formal ontology are related to upper-level ontologies in information systems (i.e., DOLCE, BFO and SUMO – see, e.g., Borgo & Vieu, 2009, for a comparison). These categories in information systems are popular but vary greatly in their relationship to categories defined in ontologies (e.g., those by *Chisholm*, *Hoffman and Rosenkrantz* or *Lowe* – see Thomasson, 2010, for an overview).

In thesaurus construction "categories" are often associated with one of the five "fundamental facets" that are attributed to S.R. Ranganathan (Aitchison et al., 2000, F2.2): personality, matter, energy, space, time. The thesaurus standard states that categories are mutually exclusive and based on shared characteristics without detailing what this precisely means. Further, it lists some examples that are "indicative but do not comprise all possible categories" (ISO 25964-1:2011, Section 5.1.2). There is also a footnote saying that "examples of high-level categories that can be used for grouping concepts into facets are: objects, materials, agents, actions, places and times" (ibid., Section 2.20). Nevertheless, categories are neither defined nor systematized in ISO 25964-1:2011.

The actual comparison of ontology and thesaurus relations was done by taking the thesaurus relationships as a basis and searched for correspondences amongst the ontological relationships. Mappings from thesauri to ontologies are useful from the perspective that there is interest to reengineer thesauri to ontologies, but unlikely the other way around. Further details about the mapping process are provided in Section 4.4.

4.2. Thesaurus relationships

Thesauri have three types of relationships: (1) equivalence, (2) hierarchical and (3) associative. We do not cover equivalence relationships in detail. **Equivalence relationships** convey that two or more synonymous or quasi-synonymous terms label the same concept (ISO 25964-1:2011, Section 2.18). One of the terms is generally chosen as the preferred term and the directedness of the equivalence relationship solely makes explicit, which single term amongst the equivalent terms is the preferred one. The preferred term is generally used to represent the concept in displays for humans. While being important for language-based access to thesauri, equivalence relationships have no structural significance in thesauri and it is for this reason that we do not include them.

There are three types of **hierarchical relationship**: (a) the generic relationship, (b) the hierarchical whole-part relationship and (c) the instance relationship. Further, hierarchical relationships "should be established between a pair of concepts when the scope of one of them falls completely within the scope of the other. It should be based on degrees or levels of superordination and subordination, where the superordinate concept represents a class or whole, and subordinate concepts refer to its members or parts" (ISO 25964-1:2011, Section 10.2.1). However, neither "class" nor "scope" is precisely defined, with the latter described as the "semantic boundary of a concept" (ibid., Section 2.50). The standard also says that for hierarchical relationships "every subordinate concept should belong to the same inherent category as its superordinate concept, i.e., both the broader and narrower term should represent a thing, or an action, or a property" (ibid., Section 10.2.1).

The *generic relationships*, referred to as *hyponymy* in terminology standards, is "the link between a class or category and its members or species" (ibid., Section 10.2.2). The role of the subordinate concept is – in linguistic terms – that of a hyponym; the one of the superordinate concept (the class/category) is that of a hypernym. For checking the correctness of applying the generic relationship, the standard recommends an all-and-some test, which is in the line of being able to say, e.g., all parrots are birds, but some birds are parrots.

Table 1

Hierarchical whole-part relationships in thesauri

Categories of relata in hierarchical whole-part relationships		Example
1st relata	2nd relata	
Systems of the body	Organs of the body	Cardiovascular system – Blood vessels – Arteries
Geographical location	Geographical location	Canada – Ontario – Ottawa
Discipline or field of discourse	Discipline or field of discourse	Science – Biology – Botany
Social entity	Social entity	Armies – Corps – Divisions

The *hierarchical whole-part relationship* is known as meronymy in the context of terminologies. The role of the subordinate concept (the part) is that of a meronym; the one of the superordinate concept (the whole) is that of a holonym. The relationship is correctly applied, if the part belongs *uniquely* to the whole, that is, there is no multihierarchy for hierarchical whole-part relations (ibid., Section 10.2.3.1). The standard expects the use of the hierarchical whole-part relationship to apply mainly to the few cases listed in Table 1. Many other linguistic uses of part-of such as those between a wheel and a bicycle need to be treated as associations in a thesaurus. The first two columns indicate restrictions on the relata using not further specified categories; the last column provides an example as it can be found in the ISO thesaurus standard.

The *instance relationship* is to be applied between a class and an instance. Again, the standard does not define these terms. 'Alps' and 'Himalayas' are provided as examples of instances for the class 'mountain regions'. In general, a thesaurus refers to instances that are given a proper name only.

Hierarchical relationships in general may be applied in a way that creates multiple top-down paths to a concept, i.e. a concept may have multiple superordinate concepts. The hierarchical relationships to the superordinate concepts may be of different types (ISO 25964-1:2011, Section 10.2.5). An example of such mixed relationship is provided in Section 4.5.

The **associative relationship** is used in "suggesting additional or alternative concepts for use in indexing or retrieval" (ISO 25964-1:2011, Section 10.3) and it is to be applied between "semantically or conceptually" related concepts that are not hierarchically related (ibid.). Further, the standard states that whenever one concept is used, "the other should always be implied within the common frames of reference shared by the users of the thesaurus. Moreover, one of the terms is often a necessary component in any explanation or definition of the other. . . . It is particularly important to establish an associative relationship between concepts that overlap in scope" (ibid.). One concept should always be "strongly implied by the other" (ibid.).

The standard lists some representative examples for associative relationships, which are summarized in Table 2. The table is organized in the same way as the one for hierarchical whole-part relations before. Notably, there is the association between an artifact and its parts. It is distinguished from the hierarchical whole-part relationships in that the part is *not* required to belong uniquely to the whole. This separation of hierarchical and associative whole-part relation is based on "separability" – a linguistic distinction rather than a logical one (Guizzardi, 2005, p. 172).

4.3. Ontological relationships in terms of their transitivity and relata

This section describes general ontological relationships by appealing firstly to the categories and relationships in philosophical literature (e.g., Lowe) and then later to the use of formal ontology in information systems (e.g., DOLCE). In Lowe's ontology (Lowe, 2005) there are orthogonal dimensions of

Table 2
Associative relationships in thesauri

Categories of relata in associative relationships		Examples
1st relata	2nd relata	
Discipline or field of study	Object or phenomenon studied	Ornithology – Birds
		Forestry – Forests
Operation or process	Agent or instrument	Crime investigation – Detectives
		Temperature control – Thermostats
Action	Action product	Weaving – Cloth
		Ploughing – Furrows
Action	Patient or Target	Harvesting – Crops
		Imprisonment – Prisoners
Object or material	Property defining Object or Material	Poisons – Toxicity
		Magnets – Ferromagnetism
Artifacts	Part of artifact	Lenses – Optical instruments
Effect	Cause	Bereavement – Death
		*Diseases – Pathogens
Object or process	Counter agent	Plants – Herbicides
		Inflammation – Anti-inflammatory agent

*Should be read: have become diseased – being infected by pathogens.

Table 3
Categories in Lowe's ontology

	Substances (things)	Properties
Universals	Substantial universals (kinds), e.g., chair[ness]	Non-substantial universals (properties), e.g., red[ness]
Particulars	Substantial individuals (objects), e.g., a specific chair	Non-substantial individuals (modes), e.g., the redness of a specific tomato

categories, universals vs. particulars, and substances/things vs. properties, between which some fundamental ontological relationships can be explained (see Table 3).

Lowe's ontology can be augmented (Bittner et al., 2004) to form the fundamental ontological relationships (see Fig. 2 and corresponding Table 4). Figure 2 displays the ontological entities as gray boxes, their relationships using connecting lines, and the relationship names along these connectors. The relationships are directed in a way shown in Table 4, which also includes whether the relationship is transitive. The naming of relationships is kept in line with the cited sources as far as possible.

The instantiation relationship (*instance-of*) is between individuals and universals, and can be applied between substantial universals and individuals as well as between their non-substantial counterparts. Properties *characterize* kinds on the level of universals, and modes *inhere-in* objects on the level of individuals.

Bittner et al. (2004) specified two relationships concerning hierarchy: the *is-a* relationship (also called subsumption) and the *part-of* relationship. For universals the is-a relation applies between kinds, and for individuals between objects.

Most of these relations are anti-transitive in that they can never be transitive. Only the is-a relationship is always transitive. Generally, the part-of relationship is considered transitive, but this is true only in terms of ground mereology, an "ontologically innocent" relationship about which nothing else is said other than it is reflexive, transitive and antisymmetric (Varzi, 2011).[10] Transitivity may not hold for

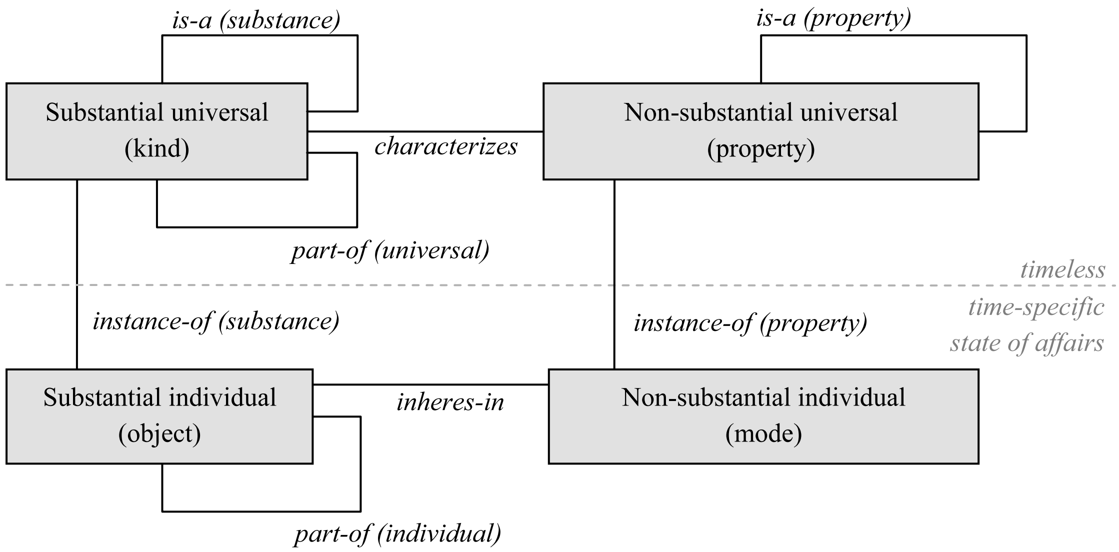

Fig. 2. Ontological relationships between fundamental relata.

Table 4
Typology of fundamental ontological relationships

Name of relationship	Categories of relata (using Lowe's categories)		Example	Transitivity
	1st relata	2nd relata		
instance-of	(1) Object	(1) Kind	(1) Specific tomato – Tomato	Anti-transit.
	(2) Mode	(2) Property	(2) Red(ness) of a specific tomato – Red(ness)	
characterizes	Property	Kind	Red(ness) – Tomato	Anti-transit.
inheres-in	Mode	Object	Red(ness) of a specific tomato – Specific tomato	Anti-transit.
is-a	(1) Kind	(1) Kind	(1) Roma tomato – Tomato	Transitive
	(2) Property	(2) Property	(2) Red(ness) – Color(edness)	
part-of	(1) Kind	(1) Kind	(1) Tomato skin – Tomato	Non-transit.
	(2) Object	(2) Object	(2) Skin of a specific tomato – Specific tomato	

mereology in general as it is used in different variants of conceptual modeling. For example, a door handle is part of the door and a door is part of the house, but the door handle would generally not be considered a part of a house (Varzi, 2007, Section 2).

To make the discussion of mereology more specific, we will use DOLCE. Specifically, Keet and Artale (2008) distinguished part-of relations based on categories in DOLCE (Gangemi et al., 2002; Keet & Artale, 2008): endurant, perdurant, quality and abstract. Endurants and perdurants are subcategories of substances/things as in Lowe's ontology explained above. Endurants are described as *wholly* present with their proper parts in the time they are present. Perdurants, instead, are "only partially present, in the sense that some of their proper temporal parts (e.g., their previous or future phases) may not be present" (Gangemi et al., 2002, Section 2.1). Endurants can change over time, keeping their identity,

[10]Sometimes proper parthood is taken as a basis for defining part-of relations. If differs from ground mereology in that it is not reflexive (Varzi, 2007, Section 2).

whilst perdurants cannot change since they have no identity. In plain words, endurant are things whilst perdurants are processes and events.

The relationships that Keet and Artale (2008) distinguished based on the DOLCE categories are summarized in Table 5, which structure corresponds to the one in Table 4. The upper half of the table grouped together transitive relationships, the lower half the non-transitive relationships. Transitivity is not possible where relata are from two different categories. Most transitive relationships connect some form of endurants, with only one involving perdurants. Relationships whose name is indented in the first column indicate that they are subtypes of the non-indented relationship. The distinction between structural-part-of and contained-in is neither well articulated nor often distinguished in the literature. Functional-part-of specializes structural-part-of by indicating that the function of the whole depends on the function of its part (Keet & Artale, 2008, pp. 99–100).

4.4. Results of the comparison of ontological and thesaurus relationships

The mapping of relationships requires a prior mapping of the categories by which the relata of the relationships are typed. Both, DOLCE's as well as Lowe's categorial system explain the categories in human language and provide examples (Gangemi et al., 2002). For capturing the meaning of the categories used in the thesaurus standard we had to rely solely on the meaning that the names of the categories convey as well as the provided examples.

Table 6 shows the result of the mapping. The first column lists thesaurus categories; the second column contains the corresponding ontology categories; the third column indicates whether we mapped to a DOLCE category or to one from Lowe. An entry showing "X" means that no match could be found and further investigation is required.

When we assigned DOLCE categories to thesaurus categories, we indicated as specific as possible DOLCE categories, but always mentioned the corresponding higher-level categories (often 'endurant') for which there is lesser doubt of having identified a correct match. Since the relata of many ontological relationships are typed through such highest-level categories, we can be confident to not have established mismatches of relationships.

Most of the thesaurus categories we identified as endurants, some as perdurants, and few as qualities. We found it difficult to match disciplines or fields of study/discourse to ontology categories at all. This will need further investigation. Where more specific categories in DOLCE were found, we traced these back to the more general category (either endurant or perdurants in these cases, as indicated in brackets). We applied the categories distinguished by Lowe in three cases: "property" corresponded to "non-substance", "class" to "universal" and "instance" to "individual".

Based on the analysis of thesaurus and ontology relationships and the mapping of categories for the relata, we can compare corresponding relationships in each.

Table 7 (summary level) and similarly structured Table 8 (focusing on hierarchical whole-part relationships) and Table 9 (focusing on associative relationships) indicate the identified matches that will be explained throughout the rest of this sub-section. The 3rd column of these tables indicates if the match can be expressed at the level of universals, individuals or both. The final column shows the transitivity of the relationships.

In a first stage of comparing thesaurus relationships to ontology relationships, we considered only the two structural relationships that the ISO thesaurus standard requires to distinguish: the hierarchical and the associative relationship. Both relationships, as we have detailed, summarize a variety of relationships, which is not a permissible practice in ontologies. Thus, both these relationships as such do not have a single corresponding relationship in ontologies and must be considered non-transitive (see Table 7).

Table 5

Classifications of part-of relationships; based on Keet and Artale (2008)

Name of relationship	Categories of relata (following DOLCE)		Example	Transitivity
	1st relata (part)	2nd relata (whole)		
structural-part-of	Endurant (1) Physical object (2) Non-physical object	Endurant (1) Physical object (2) Non-physical object	(1) Actin filament – Cell (2) Introduction – Report	Transitive
functional-part-of (function of part requires function of whole)	Endurant (functionally specified)	Endurant (functionally specified)	Car engine – Car	Transitive
spatial-part-of			*(see two options below)*	Transitive
contained-in (3D)	Endurant occupying 3D-Region	Endurant occupying 3D-Region	(Passenger) D. B. Cooper – (Specific) Aircraft	Transitive
located-in (2D)	Endurant occupying 2D-Region	Endurant occupying 2D-Region	Normandy – France	Transitive
involved-in	Perdurant	Perdurant	Paying – Purchasing	Transitive
member-of (part and whole existentially dependent)	Physical or Social object	Social object	Musician – Orchestra Sheep – Herd	Non-transit.
constitutes	Matter	Physical object	Marble – Statue	Non-transit.
sub-quantity-of	(1) Matter (same) (2) Matter (different)	(1) Matter (2) Matter	(1) 10ml wine – Wine (2) Alcohol – Wine	Non-transit.
participates-in	Endurant	Process	Enzyme – Catalytic reaction	Non-transit.

Table 6

Mapping of thesaurus categories to ontological categories

Thesaurus category	Ontology category	Category system (ont.)
Action	Process (perdurant)	DOLCE
Action product	Endurant	DOLCE
Agent	Social agent (non-physical endurant)	DOLCE
Artifacts	Endurant	DOLCE
Cause	Achievement (event, perdurant)	DOLCE
Class	Universal	Lowe
Counter agent	Agentive physical object (endurant)	DOLCE
Discipline	X	DOLCE
Effect	Accomplishment (event, perdurant)	DOLCE
Field of study/discourse	X	DOLCE
Geographical location	Spatial location (quality)	DOLCE
Instance	Individual	Lowe
Instrument	Endurant	DOLCE
Material	Endurant	DOLCE
Object	Endurant	DOLCE
Operation	Process (perdurant)	DOLCE
Part of artifact	Endurant	DOLCE
Patient	Social agent (non-physical endurant)	DOLCE
Phenomenon	Event (perdurant)	DOLCE
Process	Process (perdurant)	DOLCE
Property	Non-substance	Lowe
Social entity	Social agent (non-physical endurant)	DOLCE
Target	Physical endurant	DOLCE

Note: X indicates that we could not easily identify an appropriate match. Further investigation is required.

Table 7

Transitivity of general thesaurus relationships

Thesaurus relationship	Ontological relationship	Level	Transitivity
Hierarchical relationship	Different relationships	n/a	Non-transitive
Generic relationship (Hyponymy)	is-a	Universal	Transitive
Hierarchical part-of relationship (Meronymy)	Ground mereology*, different part-of relationships	Universal or individual	(Transitive)*
Instance relationship	instance-of	Between universal and individual	n/a
Associative relationship	Different other relationships**	n/a	Non-transitive

*Matches the transitive ground mereology only, if being a necessary (constituent) part. See also further text comments.
**There is no relation in ontologies to which associative relationships in thesauri can be mapped. See also further text comments.

The thesaurus standard gives us the option of distinguishing the three types of hierarchical relationships (the generic relationship, the hierarchical whole-part and the instance relationship) explicitly. It should be acknowledged that this finer distinction of relationships and relata is unlikely to be found in real-life thesauri. The distinctions have to be made posterior in a domain-specific thesaurus to perform an analysis as we described it here. On this basis the generic relationship corresponds to the is-a relationship in ontologies, but only if the related concepts are universals. As we noticed in the definition of the generic relationship already, it is definitely a transitive relation. Similarly, we can match the instance

Table 8

Transitivity of hierarchical whole-part relationships in thesauri

Hierarchical whole-part relationship in thesauri[a]	Ontological relation	Level	Transitivity
Relation between systems and organs of the body	structural-part-of	Universal or individual	Transitive
Relation between geographical locations	spatial-part-of	Individual	Transitive
Relation between disciplines or fields of discourse	X[b]	X[b]	(Transitive)[b]
Relation between hierarchical social structures	structural-part-of	Universal or individual	Transitive

[a]Due to a lack of a name, the thesaurus relations are referred to by indicating the categories of their relata (see Table 1).

[b]X indicates that – due to the uncertainty about the thesaurus categories (see Table 6) – we could not easily identify an appropriate match. Thus there's also uncertainty about the transitivity, which assume, though.

relationship in thesauri and the instance-of relation in ontologies. Again, we must assume that the superordinate concept fulfills the criteria of a universal and the subordinate concept (called "instance") is an individual. Transitivity does not apply for this relationship.

Both the hierarchical whole-part relationship and the associative relationship require further examination. Specifically, the general hierarchical whole-part relation, like the associative and general hierarchical relationship, faces the problem of encompassing a variety of more specific relationships. We cover each below.

Hierarchical whole-part relationships in thesauri need to be distinguished at the level of the rules that the thesaurus standard provides for the correct application of these relationships. At this level categories serve as distinguishing properties of the relations (Tables 1 and 2). These whole-part relations can be mapped to different transitive relationships as they have been defined by Keet and Artale (2008) (Table 5). Table 8 lists the results of the mapping.

In two cases we found the structural-part-of relation to be the most appropriate match. In both these cases they can be expressed at an abstract level, i.e. between universals, but also at a specific level, i.e. between individuals. This was not the case with geographical locations, which are only used to refer to specific geographical locations, i.e. applied between individuals only. Thus, the spatial-part-of relation will be at an individual level only. As a consequence, three hierarchical whole-part relationships in thesauri must be considered transitive. A special case is the relationship between disciplines and fields of discourse since we could not match these categories to any ontological category. It appears reasonable to consider them transitive, though.

The hierarchical whole-part relationship in thesauri also generally matches the ground mereology as detailed in Section 4.3. This can be concluded from the introduced rules of the ISO thesaurus standard: First, we assume that the different cases of hierarchical whole-part relations (as exemplified in Table 8) cannot be mixed with each other, since their relata must always belong to the same fundamental category. Second, we assume the different cases of whole-part relations to be *always* transitive, since the ISO standard requires the part to be always unique to the whole. The latter assumption is backed by transitivity in all of the "casually found" cases listed in Table 8. Nevertheless, transitivity does not hold, if the part is not a *necessary* part of the whole, e.g., a 'house door handle' is a "typical" part of a 'house door', but not a necessary part of it. For this reason a house door handle cannot be inferred to be part of a 'house', even though house doors are typical parts of houses (but again not necessary parts). The question of necessity will be further discussed in Section 4.5.

The finer distinctions of **associative relationships** in the rules of the thesaurus standard (Table 2) need to be referred to in order to match this relationship to ontological ones. On this basis it is difficult to see matches with ontological relationships that we included in this paper (see Table 9). This may not surprise, since the ontological relationships that we collected were either of very fundamental nature or

Table 9
Transitivity of associative thesaurus relationships

Associative relationship in thesauri[a]	Ontological relationship	Level	Transitivity
Relation between a discipline or field of study and the objects or phenomena studied	X	X	Non-transitive
Relation between an operation or process and its agent or instrument	participates-in	Universal or individual	Non-transitive
Relation between an action and the action product	X	Universal or individual	Non-transitive
Relation between an action and its patient or target	X	Universal or individual	Non-transitive
Relation between objects or materials and their defining properties	Characterizes	Universal or individual	Non-transitive
Relation between an artifact and parts of the artifact	(functional-part-of)[b]	Universal or individual	X[b]
Relation between cause and effect	X	Universal or individual	Non-transitive
Relation between an object or process and its counter agent	X	Universal or individual	Non-transitive

Notes: X indicates that there is no match amongst the limited range of ontological relationships, which we summarized in Section 4.3. There may be found formally defined relationships in other papers or custom relationships may be defined.
[a]Due to a lack of a name, the thesaurus relations are referred to by indicating the categories of their relata (see Table 2).
[b]Some of the thesaurus relations between artifacts and its parts may match the functional-part-of relation, others not. This insecurity does not allow statements about the transitivity.

focused on part-of relations. Other sets of ontological relationships that express the meaning of these associative thesaurus relations may be adopted, but also custom relations in thesauri can be defined to cover ontological relationships. These relationships would generally be non-transitive since their relata are of different categories.

Nevertheless, there are three correspondences between thesaurus associations and ontological relationships. Firstly, the association between an operation or process and an agent or instrument corresponds to the ontological (non-transitive) participates-in relation. Secondly, the associative relationship between objects or materials and their defining properties matches the ontological ‘characterizes’ relationship between kinds and properties (or objects and modes on an individual level). Similar to the instance relationship, it is thus another case where fundamental ontological things are related. Thirdly, the associative part-of relation corresponds to the functional-part-of relationship in ontology.

4.5. Pragmatic differences between thesaurus and ontology relationships

Thesaurus relationships and ontology relationships have been compared as formal modeling primitives so far, without paying attention to pragmatic aspects, i.e., how these relationships are to be applied in thesauri and ontologies. This section discusses such pragmatic aspects, which are important to consider when then the intention is to use a specific thesaurus as an ontology or vice versa.

Ontological relationships are special kinds of properties (Swoyer & Orilia, 2011, Section 1.1.5). As such *relationships are interpreted as necessary conditions in an ontology* (see also Section 3.2). The reason for this interpretation lies in the roots of ontologies in logic-based reasoning, where meaning is expressed through truth conditions (Saeed, 2009, Section 10.6).

Thesaurus relationships, in contrast, do not define properties or conditions of “concept membership”, but appear to describe a kind of *prototypical* reality with their relationships – one that describes the things in their idealized, that is, “common” and “normal” form. This makes thesauri and ontologies fundamentally different in terms of the validity and purpose of applying relationships. The correspondences between modeling entities that we identified in Section 4.4 (Tables 7–9) can thus be used as candidate mappings only and should not be interpreted as a precise recipe or algorithm sufficient for using a thesaurus as an ontology.

If a thesaurus is to be "reused" as ontology, relations in a specific thesaurus, particularly hierarchical whole-part relations, have to be validated case-by-case for their correctness as necessary condition for their respectively related universals in an ontology – a process that is generally referred to as ontological *reengineering* of thesauri, making the term "reuse" basically a misnomer in this context.

Such reengineering has to consider that relationships expressing necessary conditions in ontologies are always directed and do not imply that their inverse is true. While no attention is paid to the inverse as logical property of relationships in thesauri, thesaurus relationships are reciprocal, that is, if one concept is, e.g., hierarchically superordinate to another concept the latter is automatically subordinate under the former (ISO 25964-1:2011, Sections 10.2.1 and 14.3). Applying this to hierarchical whole-part relationships as defined in the thesaurus standard (see Table 1) means that, e.g., a 'cardiovascular system' having 'blood vessels' as parts implies that 'blood vessels' are part of a 'cardiovascular system'. Interpreted as part-of relation in ontologies the latter is a wrong statement: a blood vessel is still a blood vessel, even if it happens to not being part of the cardiovascular system at some point of time. There could also be argued that an artery is still an artery, even if some blood vessels are removed so that also former relationship would have to be constrained by a "some" quantifier in the ontology.

Further, a proper reengineering of a thesaurus along ontology lines also needs to undergo a methodical examination that may reject generic relationships in thesauri as is-a relations in ontologies. This may occur, for example, when aligning the hierarchy to a top-level ontology or when applying the OntoClean method (Guarino & Welty, 2009).

Also associative relationships may be invalid relationships in ontologies. For example, the concept 'fertilizers' in the AGROVOC thesaurus is associated with the term 'soil pollution'.[11] It has often been suggested to further refine such relationships (Sager, 1990, Section 2.2.4.2; Soergel et al., 2004) and in recent developments of the AGROVOC thesaurus this has resulted in a relationship fertilizers – *causes* – soil pollution.[12] This relationship has no value in an ontology since fertilizers do not *necessarily* cause soil pollution. Interpreting such relationships in thesauri as predicated statements in an ontology can lead to errors in reasoning.

It would also be wrong to assume that ontological relations could be translated into thesaurus relations using the correspondence tables. Ontologies are likely to make use of further formal ontology relations than we introduced here given they can express necessary conditions of universals. Such relationships may not be useful for the context and purpose that drive the application of thesaurus relationships. This also reveals that the establishment of relationships in thesauri and ontologies are differently motivated.

4.6. Discussion of the comparison results

The comparison reveals various differences and similarities between thesauri and ontologies. First of all, we recognize that most of the hierarchical relationships in the thesaurus model exist, in one form or another, as relationships in ontologies. This explains why thesauri often appear very similar to ontologies. Particularly relevant is the congruence between thesauri's *generic relationship* and ontologies' *is-a* relation as well as the parallels between thesauri's *hierarchical whole-part relation* and ontologies' *part-of* relation. These two relationships dominate typical domain modeling and also mirror each other

[11] http://aims.fao.org/en/agrovoc-term-info?mytermcode=2867.

[12] http://code.google.com/p/agrovoc-cs-workbench/ (requires login).

in thesauri and ontologies with respect to transitivity. The hierarchical whole-part relation in thesauri appears to be little different from logical mereology on formal grounds.

The need for detailing thesaurus relationships to make them comparable to ontological relationships is where the differences between thesauri and ontologies start. The fact that structural thesaurus relationships are summarized to a degree that there are only two relationships (associations and hierarchical relations) makes them highly underspecified from an ontological perspective. A consequence of this under-specification is a loss of transitivity, since the relationships can be mixed and transitivity does not hold across different types of relationships. This is not even the case if the mixed relationships are transitive by themselves. An example from the AGROVOC thesaurus[13] shall illustrate such case:

- Plant reproductive organs
  - Seed (hyponym)
    - Kernels (meronym)
      - Endosperm (meronym)
    - Testa (meronym)
  - Fruit (hyponym)

In this case generic relationships and hierarchical whole-part relations are mixed. Neither would a *testa* (seed cover) inherit any intensional properties from seeds (if it was a hyponym) nor does a *plant reproductive organ* represent a whole for a specific *testa*. It is obvious that distinguishing thesaurus relationships is necessary to express formal logical properties such as transitivity about them.

The cause for another main difference between thesauri and ontologies is the cursory usage of terms such as “class”, “instance”, “property” or “category” in the definitions of thesaurus relationships. These terms are neither defined nor put into a consistent framework. Classes, instances and properties are all considered concepts. Relationships between the concepts dissolve in the sea of general hierarchical and associative relationships. An example can be found in the terminology standard ISO 704:2009, which could be found in a thesaurus as well:

Geopolitical entity $\rightarrow$ Country $\rightarrow$ Canada

In this example, the instance relationship is mixed with the generic relationship, and thus, from an ontological perspective, the thesaurus jumps from the level of universals (country) to particulars exemplifying them (Canada). At the individual level, Canada could be further related with the spatial-part-of between individuals as indicated in Fig. 2. Another example can be taken from the AGROVOC thesaurus again where numerous specific rivers (e.g., ‘Mekong’) are listed under the concept ‘river’.[14]

Our debate of ontology relata in the first part (Section 3) showed that universals and individuals denote most fundamental entities or categories in ontological realism. They are the major cornerstones in its framework. As a result, thesauri’s hierarchical instance relationships as well as its associative relationships between objects and their properties receive a particular importance in ontologies. They cannot be considered at the same level with other relationships and be mixed with them. The opposition of universals and individuals in realism-driven ontologies pulls hierarchical thesaurus structures apart into (a) relations between universals, (b) relations between individuals and (c) instantiations between universals and individuals. Generally, just (a) is considered part of a formal ontology while (b) and (c) are concerned with the description of a specific state of affairs at some point of time. A similar case

[13]http://aims.fao.org/en/pages/594/sub?mytermcode=27986.

[14]http://aims.fao.org/en/pages/594/sub?mytermcode=6617.

is the importance of distinguishing the category of properties from substances/things as it was laid out through Lowe's categorization in the beginning of Section 4.

A final difference between thesauri and ontologies in the row of imprecisely defined entities lies in the notion of "categories". In ontologies categories are used with significant meaning and it is considered an important question, what kind of foundational ontology one is committed to. While the role of categories is seen critical in connection with ontologies (Lenat, 2005, Section 3; Merrill, 2010, Section 5; Thomasson, 2010, Section 1.5), an "intuitive" application and interpretation as it is suggested in the thesaurus standard is certainly not helpful in making a difference.

5. Conclusions

We started our paper with the question, what the differences and commonalities between thesauri and ontologies are. For this purpose we compared thesauri and ontologies by mapping their relationships and relata. We put particular attention to the transitivity of relationships, but also considered pragmatic aspects distinguishing the way relationships are applied and the meaning of relata is expressed in thesauri and ontologies. Finally, we summarized, whether or not ontologies are better than thesauri for typical purposes of thesauri.

We could show that thesauri in their theoretical definition, through ISO 25964-1:2011, have many similarities with ontologies in the way they define structural relationships, particularly with respect to the is-a relation in ontologies and the generic relationship in thesauri. However, thesauri have significant deficiencies from the perspective of ontological realism, most importantly an indifferent agglomeration of ontologically different entities and relationships: thesaurus concepts can correspond to universals, but also represent a single individual, pure collections or mixed collections in ontologies; hierarchical relationship in thesauri may correspond to different kinds of part-of relationships, the is-a relationship or the instance-of relationship in an ontology. While hierarchical thesaurus relationships may generally be transitive, the possible mixing of different kinds of relationships may cause "breaks" in the hierarchical transitivity.

Thesauri and ontologies also differ in the way relationships are applied when modeling specific things in a domain. While ontology relationships express necessary membership conditions for their relata, thesaurus relationships are established out of other motivations, most likely based on prototypical properties of the things that the relata represent. As a result ontology relationships do not imply the inverse relationship between two relata to be true (their "reciprocity") as it would generally be assumed for thesaurus relationships.

The different principles behind applying relationships also makes obvious that thesauri cannot be considered a less expressive type of ontology, but that thesauri and ontologies are simply different models with just apparently similar structures. This makes it also necessary to investigate in depth the competitive usefulness of thesauri, ontologies and other types of models for different application scenarios in order to guide practitioners in the adequate choice between them. Publications claiming that ontologies have manifold usages such as indexing – a stereotypical usage of thesauri – or text analysis (Stevens & Lord, 2009) either do not back these claims through examples or show little evidence that the examples are truly ontologies as we discussed them in this paper. Further research is required to substantiate the usefulness of ontologies for applications other than specifying and classifying things in reality and reasoning about these things. The different ways of conveying meaning as well as the (in-)transitivity of relationships must be expected to have significant influence on the usefulness of thesauri and ontologies.

Instead of using either a thesaurus or an ontology, it may also be that applications benefit from applying them in parallel or in conjunction with other models such as terminologies providing definitions. There has already been observed a trend towards providing definitions in thesauri (Aitchison et al., 2000, Section D5.2.1). Definitions in natural language are also useful in ontologies, because providing necessary or essential properties may often not be feasible in practical ontology construction and non-ontological forms of definitions may have to be used complimentary. Models may also positively influence each other: the is-a hierarchy of ontologies may provide more stable models for generic relationships in thesauri. This, in turn, may improve terminological term definitions since many of them start in way of saying “X is a Y that…”, e.g., “infants are humans that…”.

The identified differences between thesauri and ontologies must also be taken into account when aligning or mapping thesauri and ontologies, when reusing thesauri as ontologies or when reengineering thesauri to become ontologies. Simple mappings and alignments between thesauri and ontologies based on their labels or names are unlikely to provide useful results related to reality. The example of ‘domestic wastewater’ (Section 3.4) has shown that a thesaurus may often summarize a set of different intensions in concepts. Applying relations between such vaguely defined relata for logical reasoning – just in the same way as with universals – are likely to provide results that are in contradiction with reality – no matter whether these relations are formally well-defined or not.

The need to mediate between the different ways of expressing meaning in thesauri and ontologies appears to be often ignored when using structured vocabularies as ontologies. An example is the Ontology Alignment Evaluation Initiative (OAEI),[15] which claims to deal with schema matching and ontology integration and often holds workshops on ontology matching[16] in conjunction with the International Semantic Web Conference.[17] It is not transparent based on which principles matches are established. Further, the test cases are often thesauri, e.g., the NCI Thesaurus, but not ontologies at all. While mappings are common for thesauri and other types of vocabularies (BS 8723-4:2007, 2007), such approaches are not the same as ontology matching, because the result is not useful for logic-based reasoning as it became obvious through the comparison made here.

In future studies we plan to reengineer a specific thesaurus into an ontology and thereby develop a general method for reengineering thesauri into ontologies (Kless et al., 2012). The comparison we presented here provides a rigorous start for such reengineering.

Acknowledgements

We thank Jutta Lindenthal and the anonymous reviewers for their feedback to earlier drafts of this paper.

References

Aitchison, J. & Clarke, S.D. (2004). The thesaurus: a historical viewpoint, with a look to the future. In *The Thesaurus: Review, Renaissance and Revision* (pp. 5–21). New York: Haworth Press.

Aitchison, J., Gilchrist, A. & Bawden, D. (2000). *Thesaurus Construction and Use: A Practical Manual* (4th ed.). London: Aslib.

[15] http://oaei.ontologymatching.org/.

[16] http://om2011.ontologymatching.org/.

[17] http://iswc2011.semanticweb.org/.

ANSI/NISO Z39.19-2005 (2005). *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*. Bethesda, MD: NISO Press.

Armstrong, D.M. (1978). *A Theory of Universals: Volume 2: Universals and Scientific Realism* (Vol. 2). Cambridge: Cambridge Univ. Press.

Bittner, T. (2009). Logical properties of foundational mereogeometrical relations in bio-ontologies. *Applied Ontology*, *4*(2), 109–138.

Bittner, T. & Donnelly, M. (2007). Logical properties of foundational relations in bio-ontologies. *Artificial Intelligence in Medicine*, *39*(3), 197–216.

Bittner, T., Donnelly, M. & Smith, B. (2004). Individuals, universals, collections: on the foundational relations of ontology. In *Proceedings of the 3rd International Conference on Formal Ontology In Information Systems* (pp. 37–48). Amsterdam: IOS Press.

Borgo, S. & Vieu, L. (2009). Artefacts in formal ontology. In *Philosophy of Technology and Engineering Sciences* (pp. 273–307). Amsterdam: North-Holland.

BS 8723-3:2007 (2007). Structured vocabularies for information retrieval. Guide. Vocabularies other than thesauri. British Standards Institution (BSI), Committee IDT/2/2.

BS 8723-4:2007 (2007). Structured vocabularies for information retrieval. Guide. Interoperation between vocabularies and other components of information storage and retrieval systems. British Standards Institution (BSI), Committee IDT/2/2.

Ceusters, W., Smith, B. & Goldberg, L. (2005). A terminological and ontological analysis of the NCI Thesaurus. *Methods of Information in Medicine*, *44*(4), 498.

Chowdhury, G.G. & Chowdhury, S. (2007). *Organizing Information: From the Shelf to the Web*. London, UK: Facet Publishing.

DIN1463-1:1987 (1987). Erstellung und Weiterentwicklung von Thesauri: Einsprachige Thesauri (German Standard No. DIN 1463-1). The German Institute for Standardization.

Gangemi, A., Guarino, N., Masolo, C., Oltramari, A. & Schneider, L. (2002). Sweetening ontologies with DOLCE. In *Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web*. Lecture Notes in Computer Science (Vol. 2473, pp. 223–233). Berlin/Heidelberg: Springer.

Gilchrist, A. (2003). Thesauri, taxonomies and ontologies – an etymological note. *Journal of Documentation*, *59*(1), 7–18.

Gruninger, M., Bodenreider, O., Olken, F., Obrst, L. & Yim, P. (2008). Ontology Summit 2007 – Ontology, taxonomy, folksonomy: understanding the distinctions. *Applied Ontology*, *3*(3), 191–200.

Guarino, N., Oberle, D. & Staab, S. (2009). What is an Ontology? In S. Staab and R. Studer (Eds.), *Handbook on Ontologies*. International Handbooks on Information Systems (2nd ed., pp. 1–17). Heidelberg: Springer.

Guarino, N. & Welty, C.A. (2009). An overview of OntoClean. In *Handbook on Ontologies* (2nd ed., pp. 201–220). Heidelberg: Springer.

Guizzardi, G. (2005). Ontological foundations for structural conceptual models. PhD thesis, University of Twente, Enschede, The Netherlands.

Hodge, G. (2009). *Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files*. Washington, DC: The Digital Library Federation, Council on Library and Information Resources.

ISO 1087-1:2000 (2000). Vocabulary – Part 1: Theory and application (International Standard No. BS ISO 1087-1:2000). Terminology work. International Organization for Standardization.

ISO 25964-1:2011 (2011). Information and documentation – Thesauri and interoperability with other vocabularies – Part 1: Thesauri for information retrieval (International Standard No. ISO 25964-1). International Organization for Standardization.

ISO 2788:1974 (1974). Documentation – Guidelines for the establishment and development of monolingual thesauri (International Standard No. ISO 2788). International Organization for Standardization.

ISO 2788:1986 (1986). Documentation – Guidelines for the establishment and development of monolingual thesauri (International Standard No. ISO 2788). International Organization for Standardization.

ISO 704:2009 (2009). Principles and methods (International Standard No. ISO 704). Terminology work. International Organization for Standardization.

ISO 860:2007 (2007). Harmonization of concepts and terms (International Standard No. ISO 860). Terminology work. International Organization for Standardization.

Keet, C.M. & Artale, A. (2008). Representing and reasoning over a taxonomy of part-whole relations. *Applied Ontology*, *3*, 91–110.

Kless, D., Jansen, L., Lindenthal, J. & Wiebensohn, J. (2012). A method for re-engineering a thesaurus into an ontology. In *Proceedings of the 7th International Conference on Formal Ontology in Information Systems*, Graz, Austria. Amsterdam: IOS Press.

Kuśnierczyk, W. (2006). Nontological engineering. In *Proceedings of the 4th International Conference on Formal Ontology in Information System* (pp. 39–50). Amsterdam: IOS Press.

Lenat, D.B. (2005). Applied ontology issues. *Applied Ontology*, *1*, 9–12.

Lowe, E.J. (2005). *The Four-Category Ontology*. Oxford: Oxford Univ. Press.

MacLeod, M.C. & Rubenstein, E.M. (2005). Universals. In J. Fieser and B. Dowden (Eds.), *Internet Encyclopedia of Philosophy*. Available at http://www.iep.utm.edu/universa/.
Merrill, G.H. (2010). Realism and reference ontologies: considerations, reflections and problems. *Applied Ontology*, *5*(3,4), 189–221.
Munn, K. & Smith, B. (Eds.) (2009). *Applied Ontology. An Introduction*. Frankfurt/Lancaster: Ontos Verlag.
Obrst, L. (2006). *The Ontology Spectrum and Semantic Models*. MITRE Information Semantics Group, Information Discovery and Understanding, Center for Innovative Computing and Informatics.
Oltramari, A., Gangemi, A., Guarino, N. & Masolo, C. (2002). Restructuring WordNet's top-level: the OntoClean approach. In *LREC2002*, Las Palmas, Spain.
OWL2-Overview (2009). *OWL 2 Web Ontology Language. Document Overview*. World Wide Web Consortium (W3C). Available at: http://www.w3.org/TR/owl2-overview/.
Poli, R. & Obrst, L. (2010). The interplay between ontology as categorial analysis and ontology as technology. In R. Poli, M. Healy and A. Kameas (Eds.), *Theory and Applications of Ontology: Computer Applications* (pp. 1–26). Dordrecht: Springer.
Saeed, J.I. (2009). *Semantics* (3rd ed.). Chichester: Wiley-Blackwell.
Sager, J.C. (1990). *A Practical Course in Terminology Processing*. Philadelphia, PA: John Benjamins Publishing.
Schulz, S., Schober, D., Tudose, I. & Stenzhorn, H. (2010). The pitfalls of thesaurus ontologization – the case of the NCI thesaurus. In *Proceedings of the AMIA Annual Symposium* (Vol. 2010, p. 727). Bethesda, MD: Amer. Med. Inform. Assoc.
Seibt, J. (2010). Particulars. In R. Poli and J. Seibt (Eds.), *Theory and Applications of Ontology: Philosophical Perspectives* (pp. 23–55). Dordrecht: Springer.
SKOS-Reference (2009). *SKOS Simple Knowledge Organization System. Reference*. World Wide Web Consortium (W3C). Available at: http://www.w3.org/TR/skos-reference/.
Smith, B. (1989). Logic and formal ontology. In J.N. Mohanty and W. McKenna (Eds.), *Husserl's Phenomenology: A Textbook* (revised version as available online, pp. 29–67). Lanham, MD: University Press of America. Available at: http://ontology.buffalo.edu/smith/articles/lfo.html.
Smith, B. (2004). Beyond concepts: ontology as reality representation. In A. Varzi and L. Vieu (Eds.), *Proceedings of the International Conference on Formal Ontology and Information Systems* (pp. 73–84). Amsterdam: IOS Press.
Smith, B. & Ceusters, W. (2010). Ontological realism: a methodology for coordinated evolution of scientific ontologies. *Applied Ontology*, *5*(3,4), 139–188.
Smith, B., Ceusters, W. & Temmerman, R. (2005). Wüsteria. *Studies in Health Technology and Informatics*, *116*, 647–652.
Smith, B., Kuśnierczyk, W., Schober, D. & Ceusters, W. (2006). Towards a reference terminology for ontology research and development in the biomedical domain. *Proceedings of KR-MED*, Baltimore MD, USA (Vol. 2006, pp. 57–65). Citeseer.
Soergel, D., Lauser, B., Liang, A., Fisseha, F., Keizer, J. & Katz, S. (2004). Reengineering thesauri for new applications: the AGROVOC example. *Journal of Digital Information*, *4*(4).
Stevens, R. & Lord, P. (2009). Application of ontologies in bioinformatics. In *Handbook on Ontologies* (2nd ed., pp. 735–756). Berlin: Springer.
Stock, W.G. (2010). Concepts and semantic relations in information science. *Journal of the American Society for Information Science and Technology*, *61*(10), 1951–1969.
Swoyer, C. & Orilia, F. (2011). Properties. In E.N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Available at: http://plato.stanford.edu/archives/fall2011/entries/properties/.
Thomasson, A. (2010). Categories. In E.N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Available at: http://plato.stanford.edu/archives/fall2010/entries/categories/.
Uschold, M. & Grüninger, M. (2004). Ontologies and semantics for seamless connectivity. *ACM SIGMOD Record*, *33*(4), 58–64.
van Rees, R. (2003). Clarity in the usage of the terms ontology, taxonomy and classification. In *Construction Informatics Digital Library* (paper w78-2003-432). Ljubljana University.
Varzi, A.C. (2007). Spatial reasoning and ontology: parts, wholes, and locations. In M. Aiello, I. Pratt-Hartmann and J. Benthem (Eds.), *Handbook of Spatial Logics* (pp. 945–1038). Dordrecht: Springer.
Varzi, A.C. (2011). Mereology. In E.N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Available at: http://plato.stanford.edu/archives/spr2011/entries/mereology/.
Woozley, A.D. (1967, 2004). Universals, a historical survey. In E. Craig (Ed.), *Routledge Encyclopedia of Philosophy*. London: Routledge.