An automatic method for reporting the quality of thesauri

Javier Lacasta[a,*], Gilles Falquet[b], F. Javier Zarazaga-Soria[a], Javier Nogueras-Iso[a]

[a]*Computer Science and Systems Engineering Dept., Universidad de Zaragoza, Spain*
[b]*Centre Universitaire d'Informatique, Université de Genève, Switzerland*

Abstract

Thesauri are knowledge models commonly used for information classification and retrieval whose structure is defined by standards such as the ISO 25964. However, when creators do not correctly follow the specifications, they construct models with inadequate concepts or relations that provide a limited usability. This paper describes a process that automatically analyses the thesaurus properties and relations with respect to ISO 25964 specification, and suggests the correction of potential problems. It performs a lexical and syntactic analysis of the concept labels, and a structural and semantic analysis of the relations. The process has been tested with Urbamet and Gemet thesauri and the results have been analysed to determine how well the proposed process works.

Keywords: Thesaurus, Digital libraries, Information retrieval, Thesaurus quality, Ontology alignment

1. Introduction

In information retrieval systems (IR), resources are frequently classified using thesauri or other simple knowledge models. The main reasons for this generalized use are their simple structure, the existence of established standards [1] and the integrated support provided by most catalog tools. The use

---

*Corresponding author
Email addresses:* `jlacasta@unizar.es` (Javier Lacasta),
`Gilles.Falquet@unige.ch` (Gilles Falquet), `javy@unizar.es` (F. Javier Zarazaga-Soria), `jnog@unizar.es` (Javier Nogueras-Iso)

of fully formalized knowledge structures (such as ontologies) is an alternative for these models, but they are more difficult to create and maintain and, in many contexts, there is no a real need of the additional functionality they provide.

The construction of a thesaurus in an area of knowledge requires a careful selection of the concepts and their interrelations in an appropriate general to specific hierarchy [2]. Standards such as ISO 25964 [1] describe the main features the concepts and relations must have. However, the lack of experience in their creation, the time and costs savings, or the over-adaptation to a data collection produce models with heterogeneous concepts and relations [3]. It is relatively common to find thesauri with terminological heterogeneity, overload of specificity, or even lexical issues in concept labels. Additionally, since their relations are too generic, they usually contain unclear hierarchies that are difficult to interpret [4, 3]. For example, in the Urbamet thesaurus [5] *car* and *vehicle* are in different branches. Whereas the first concept is considered as a "mean of transport", the second one is considered as a part of the "transportation management". In this context it would be better if *car* were the narrower term of *vehicle* in the "mean of transport" branch. Additionally, in the "transportation management" branch, the *vehicle* concept could be replaced with *vehicle management* or a similar concept.

Many thesauri issues become irrelevant in their original context, because users are accustomed to them. However, they may be a problem for other organizations and casual users that want to reuse these thesauri in their applications, or to use them as a starting point for building more formal knowledge models (e.g., formal ontologies). In these cases of reuse, the quality of thesauri must be evaluated so that the defects can be corrected.

This paper proposes a process that advances in the detection of the syntactic and semantic quality with respect to ISO 25964 specification [1]. It performs a lexical and syntactic analysis of the concept labels, and a structural and semantic analysis of the relations. Lexical and syntactic analysis allow identifying description issues such as the use of acronyms in preferred labels or the detection of concepts that describe too complex ideas. Structural analysis focuses on detecting sets of incompatible relations (such as cycles) and mandatory properties that are not provided. Finally, semantic analysis studies the suitability of the broader/narrower (BT/NT) relations. This semantic analysis is done by looking for compatible relations in Word-Net [6] and DOLCE [7] ontologies. The alignment of thesauri with formal models has been used in the last years to increase thesaurus semantics. Since

ontologies provide a more precise definition of their concepts and relations, they can be used to replace some items (concepts or relations) of a thesaurus with more precise ones. In this paper, we describe how to use this approach to analyse the quality of thesaurus relations. The proposed process works with both monolingual and multilingual thesauri, but some of the steps used to perform the alignment between thesauri and WordNet/Dolce take advantage of multilingual labels of concepts to improve the matching.

The paper is structured as follows. Section 2 describes the concept of quality in thesauri and the features that are relevant to measure it. Section 3 introduces the proposed analysis method. Section 4 shows the quality analysis of Urbamet and Gemet thesauri. Section 5 reviews other approaches related to thesauri quality detection and compares them with our approach. The paper ends with a discussion about the process, some conclusions, and an outlook on future work.

2. Background in thesaurus quality measures

According to ISO 8402 [8], "the quality" is a measure of excellence or a state of being free from defects, deficiencies and significant variations. It defines the quality as "the totality of features and characteristics of a product or service that bears its ability to satisfy stated or implied needs".

The main sources to identify the quality features of a thesaurus are the existent construction guidelines. They range from practice manuals such as Aitchison et al. [9], to the current international standard ISO 25964 [1]. Kless and Milton [10] aggregate the quality notions in thesaurus literature and they describe a range of abstract measurement constructs, which allow an empirically testable evaluation. It classifies these quality measures according to the different parts of the thesaurus they affect: concepts, terms, structure and documentation parts, and as a whole. These measures include some that can be transformed into rules to be interpreted in an automatic way, such as the number of words defining the number of concepts, or the degree to which there are no redundant terms in the thesauri. However, most of them, such as the degree in which the concepts are in the scope of the thesaurus or the proportion of relevant concepts of the field that are covered, are too general and abstract to be automated. They can be considered more as a general guide of the aspects that are important in a thesaurus than as specific features that need to be reviewed. Somehow, Pinto [11] obtains similar results through a survey with students, young researchers, librarians and experts

in information. The result is a set of quantitative measures indicating the perceived importance of the different aspects of thesaurus quality (structural, functional, formal, and external). As in the previous work, most of the measures are guidelines of general aspects to review, but in this case they are focused on the perception of the user (e.g., perceived thesaurus structure or perceived performance). Finally, Mader and Haslhofer [12] focus on analysing the relevance of technical aspects for quality, such as the lack of definitions or the existence of cycles in the hierarchical relations (it also includes other features, such as the use as Linked Data, which are outside the thesaurus scope). Then, the importance of these features for the quality of a thesaurus is established through a survey with vocabulary managers, term contributors and users of thesauri.

In this paper, to determine the quality of a thesaurus, we have selected the following features:

Property completeness measures: These measures are focused on the identification of lacking properties. We analyse the completeness and uniqueness of preferred labels and completeness of definitions.

Property content measures: Their objective is to locate invalid values inside labels. We focus on detecting non-alphabetic characters, adverbs, initial articles, and acronyms (in preferred labels).

Property context measures: These are focused on identifying anomalies involving several labels. This includes detecting duplicated labels and inconsistencies in the use of uppercase and plurals.

Property complexity measures: They provide a measure of the syntactic complexity of the labels, in terms of the use of prepositions, conjunctions and adjectives.

Relation coherence measures: They indicate if the relations are complete, coherent, and semantically correct. RT analysis focuses on detecting non-informative relations (they link hierarchically related concepts). BT/NT analysis searches for cycles in the model, unlinked concepts and relations that do not associate a superordinate with a subordinate concept. According to ISO 25964, the superordinate must represent a class or whole and subordinate its members or parts.

They are technical features obtained from the rules and recommendations in ISO 25964 standard. These rules include the same thesaurus related features as Mader and Haslhofer [12] and additional ones related to the syntaxis and semantics of the labels, concepts and relations (e.g., the use of adverbs, acronyms, and the meaning of BT/NT relations).

3. Thesaurus validation process

This section describes the proposed process to analyse the quality of thesauri and to detect issues according to the selected features. Figure 1 shows the main components of the process. The property quality analysis performs all the property related checks, and the relation quality analysis performs those analysis involving relations.
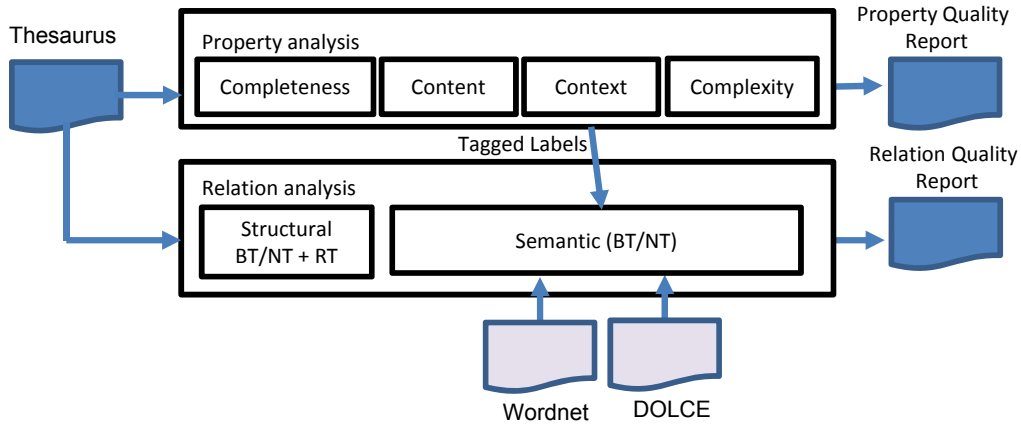

Figure 1: Thesaurus validation process

The property analysis component detects missing properties in the concepts, and labels with incorrect structure. These property checks are mainly based on lexical and syntactical analysis and involve well known natural processing language techniques, such as stemming and POS tagging. The relation analysis component starts analysing the omission or excess of relations from a structural point of view (e.g., BT/NT cycles). Then, it performs a semantic analysis of BT/NT relations to identify relations whose meaning is not compatible with the BT/NT specification. Following subsections describe each of these steps in detail.

3.1. Property analysis

This step verifies the quality of the main properties of the concepts (labels and definitions). It checks the completeness, content, context and complexity of these elements.

The **property completeness analysis** measures show the degree in which the mandatory properties are provided. That is, the *completeness and uniqueness of preferred labels* and the *completeness of definitions*. In the case of preferred labels, quality is calculated as the percentage of concepts with a single preferred label per language (there may be zero or more than one). Concerning definitions, quality is measured as the percentage of provided definitions with respect to the needed ones (1 per concept and language). Obtaining these measures is trivial as we only need to check the existence or non-existence of the required properties.

The **property content analysis** detects isolated anomalies in the labels. We check the *existence of non-alphabetic characters in the labels*, the *use of adverbs / initial articles in the labels*, and the *use of acronyms in the preferred labels*. Their quality is measured in all cases as the percentage of labels from the total that fulfil each condition. Non-alphabetic characters and acronyms are checked using regular expressions. The location of adverbs and initial articles is done using a POS tagger adequate to the label language. The tags they produce mark nouns, adjectives, prepositions, adverbs, etc. Having these POS tags, the detection of adverbs, initial articles, and other features described latter in this section is direct.

The **property context analysis** measures the anomalies that involve several labels. They look for the *existence of duplicated labels*, the *consistent use of uppercase in the labels*, and the *consistent use of plurals in the labels*. The duplicated labels feature quality is measured as the number of different labels with respect to the total number of labels. They are detected directly by string comparison. With respect to the consistent use of uppercase and plurals, the quality is measured as the ratio of labels starting with *lowercase/have a plural core* with respect to the total number of labels. Since plural convention changes depending on the language, we search for its consistent use, so the most common is considered the correct one (we deal with English, Spanish, and French). To determine plural forms, we do not analyse the label termination, but the noun nucleus after removing prepositional components. For example, the Spanish "plano de bloques" term (Block plan) is a plural prepositional phrase, but the noun "plano" is singular. Prepositional phrases are detected using the previous POS tagger. With

this approach, it is possible that not all the elements tagged as problematic are really erroneous, because the plural choice is acceptable when it is used to distinguish between countable and uncountable meanings of a noun. For example, to distinguish "painting" as a process and "paintings" as a collection of pictures. However, since this can lead to confusion, we think that it is better to tag these elements as erroneous and suggest the use of qualifiers. The detection of initial upper cases is done with regular expressions, and plural identification with a version of the Solr minimal stemmer [13].

The **property complexity analysis** measures if the labels are too complex, because it probably indicates that the associated concept is inadequate for classification. ISO 25964 recommends the use of noun phrases instead of prepositional phrases in English, and it also indicates that too complex concepts must be divided into the parts that compose them (e.g., those that use conjunctions). Therefore we focus on the *detection of prepositional phrases in the English labels*, and the *use of noun phrases too long or with conjunctions*. In general, simpler English prepositional phrases can be replaced with adjectives + substantive. When this is not possible, it usually means that the phases describe concepts too complex and too specific to be useful for classification. In Spanish or French, prepositional phrases cannot be avoided, so they have not been checked. With respect to adjectives, we consider that a label with more than three adjectives or with conjunctions should be reviewed, as it describes a concept too specific or complex to be useful. The quality of these features is measured as the percentage of labels that fulfil the indicated conditions. They are detected using the same a POS tagger used to detect adverbs and articles, but focused on detecting prepositions, conjunctions and adjectives.

3.2. Relation analysis

The **relation coherence analysis** focuses on validating the structure and semantics of the relations. The structure analysis is simple as it only requires identifying if *RT relations are informative*, if there are *orphan concepts*, and if there are *BT/NT cycles*. The RT relations quality is measured by the percentage of RT relations that do not involve a hierarchically related concept: these relations are considered informative. The connection degree of the thesaurus is measured as the percentage of concepts that are top concepts or have a broader term (they are not orphan). Cycles are located using a modified version of Tarjan's strongly connected components algorithm [14] that identifies the relation that generates the cycle (it points

out to a broader concept). The quality of this feature is measured as the percentage of concepts in the thesaurus not involved in a cycle.

The last measure calculates the *semantic correctness of BT/NTs*. It is measured as the percentage of BT/NT relations that relate a superordinate concept with a subordinate one according to ISO 25964. To determine if the meaning of the BT/NT relations is correct, it matches the thesaurus concepts with WordNet lexical database in a first step, and DOLCE ontology [7] in a second step. Then, it compares the original relations in the thesaurus with those provided in WordNet and DOLCE for the equivalent concepts. If the equivalent relation is not compatible with the BT/NT meaning, the original relation in the thesaurus is considered erroneous.

WordNet is a lexical database originally in English that groups nouns, verbs, adjectives and adverbs into sets of cognitive synonyms (synsets), each expressing a distinct concept. Synsets are interlinked by means of conceptual-semantic and lexical relations providing a hypernym/hyponym hierarchy of semantically related concepts. It is ideal for the quality analysis task as it provides conceptual-semantic and lexical relations between general and thematic concepts, such as hypernym and meronym hierarchies that can correspond with "generic" and "whole-part" specializations of BT/NT relations. However, WordNet terminology is focused on a specificity level that do not usually fit with the one provided by a thematic thesauri. In this context, DOLCE ontology can be useful as it provides very abstract categories and relations that can be used to analyse relations independently of the thesaurus domain. It provides top level categories of concepts with a deep semantic net of relations between them. Some of these relations, such as "participant'" or "exact location of" intrinsically provide a superordinate and subordinate meaning and they can be considered as BT/NT specializations, but others do not have a compatible meaning. In our process, we have used WordNet and DOLCE, but they could be replaced with other ontologies with the same characteristics (coverage of the vocabulary similar to WordNet and a formal top level ontology with a very rich relation structure such as DOLCE).

In this approach, we do not deal with the "instance" relation defined in ISO 25964. The subordinate of an instance relation is usually a proper noun, which identifies an object of a given class (the superordinate). However, WordNet only contains a few proper nouns in very specific contexts, such as general administrative divisions, geographical entities, and important authorities (DOLCE has none). This precludes the alignment with WordNet or DOLCE using the same techniques used for the rest of the concepts.

Fortunately, outside the indicated fields, thematic thesauri rarely include instances, and therefore, ignoring them is not a serious limitation.

3.2.1. Alignment with WordNet and DOLCE

Our process to align a thesaurus and WordNet is based on the work of Lacasta et al. [15]. This work aligns the main noun contained in the preferred English label of each concept with WordNet to obtain "broader match" alignments. To disambiguate polysemic terms, it uses monosemic terms and the abstracts of a collection classified with the thesaurus as context. This approach is specially useful to align noun-based labels, such as those in a thematic thesaurus, and WordNet nouns. However, the quality of the resulting alignments is not good enough for the task of detecting problems in the thesaurus relations. To refine BT/NT relations, we need both of the concepts involved in the relation to be correctly aligned. Each unaligned (or wrongly aligned) concept produces that none of the relations in which it participates are correctly analysed. Due to the tree structure of the thesaurus, a few errors in the alignment of inner concepts (not leaves) can greatly reduce the quality of the relation analysis.

To improve the alignment performance, we have extended here the work of Lacasta et al. [15] to take into account multilingual labels of concepts (current implementation deals with English, Spanish, and French) instead of using only the English ones. Since all the labels of a thesaurus concept should have an equivalent meaning (independently of the languages), all of them can be used in the alignment process to identify the concept meaning. The official WordNet is only in English, but it has been extended by different projects to incorporate additional languages. From them, we use the Open Multilingual Wordnet (OMWordNet)[1], which aggregates many of these works. This approach has two main advantages. On the one hand, the use of additional labels implies that the probability of finding a match increases. If labels in one language do not provide a match, there are others that can provide them. On the other hand, if facilitates the disambiguation of polysemic terms. It is not very common to find exact synonyms between languages (i.e. terms that have the same sets of meanings). Terms usually only share a subset of their meanings. This characteristic is useful for disambiguation, as it can be considered that the meaning of a concept is the intersection of the

---

[1]http://compling.hss.ntu.edu.sg/omw/

9

meanings of the labels that describe it. Additionally, to disambiguate poly-
semic words, we use the whole set of already aligned concepts as a context
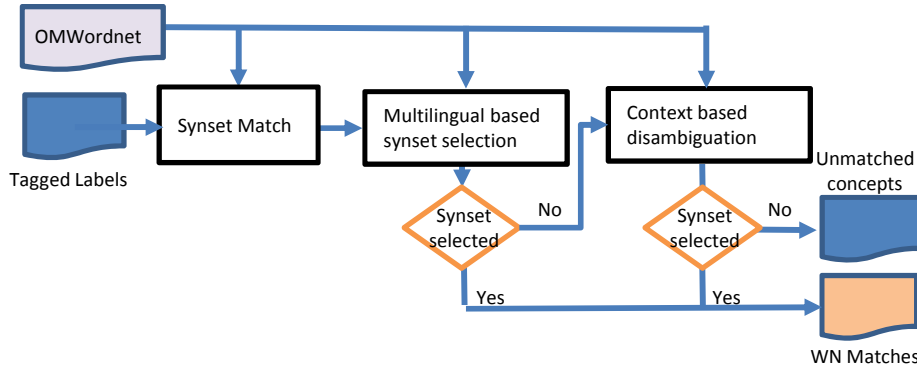network in which the new concept has to fit.

Figure 2: Synset selection process

Figure 2 describes the alignment steps in detail. This alignment generates
a collection of possible WordNet senses (synsets) for each thesaurus concept
that is then processed to identify the closest one.

The first step of the process is to match each label with OMWordNet. It
performs a multi-stage match (see Algorithm 1) to each label and its singular
version (obtained with a plural stemmer). An alternative to plural stemming
could be lemmatization. It reduces the terms to their lemma, and it is quite
used in information retrieval to find terms similar to those in the queries.
However, this approach is not suitable in this context. In any language,
there are many different terms sharing the same lemma (e.g., *savings* and
save have as lemma *save*). Therefore, if terms are lemmatized ,there are more
possible senses to select in WordNet, making disambiguation more difficult.
As plural stemming is more conservative and it introduces less ambiguity,
the alignment is simpler.

Initially, the process looks for an exact match. It there is no equivalence,
the POS tags from the concept analysis step are used to identify prepositional
phrases and reduce them to noun phrases. Then, a new alignment is done.
If there is still no alignment, adjectives are removed one by one from general
to specific and then matched. This allows identifying the closest possible
match in OMWordNet as the one that requires the removal of fewer adjectives
(it is a kind of world edit distance, but using only the removal operation).
For example, WordNet does not contain "city master plan", but it contains

"master plan" and "plan". Both alignments are valid, but "master plan" is obtained with the removal of just one adjective and "plan" requires the removal of two, so the first one is selected. If none of the concept labels is aligned, the concept is tagged as non-alignable.

function getWordNetSense.
input : label *//label to align = a pair (string, language)*
output : senses *//list of senses*
             processLevel *//level of processing performed to the labels*

senses ← getExactWordNetMatch(label);
processLevel ← 0;
if *senses = ∅* **then**
    label ← removePlural(label);
    senses ← getExactWordNetMatch(label);
    processLevel ← 1;
    if *senses = ∅* **then**
        label ← removePrepositionalPhrase(label);
        senses ← getExactWordNetMatch(label);
        processLevel ← 2;
        while *senses = ∅ & hasAdjective(label)* **do**
            label ← removeMostSpecificAdjective(label);
            senses ← getExactWordNetMatch(label);
            processLevel ← 3;
        end
    end
end
return senses, processLevel;

Algorithm 1: Alignment of label to WordNet

The result of the label alignment is a pool of synsets associated to each concept label with an accuracy measure, which indicates the degree of processing that has been required to obtain it (original label, without prepositions, without X adjectives). We also tag the matches obtained from the original label as "exact matches", and the rest as "broader matches".

Since many nouns are polysemic, it is common that each original label generates several possible senses. In this context, the pool of synsets is analysed to select the one that represent best the meaning of the concept (see Algorithm 2). First, we select the best quality synsets of each language labels (obtained from less processed labels). Then, we calculate a consensus-based intersection between languages to extract the synsets that are common to most languages (at least 2 of them), and have an alignment of the best possible quality between those available in at least one language. In this context, the intersection between the identified synsets represent the core meaning

11

of the concept. When no synset is shared in at least two languages, all the synsets are expanded using WordNet hypernyms hierarchy. The new synsets are then added to the pool with the alignment quality of their hyponyms and the intersection of synsets per language is repeated again. In this phase, we select less specific senses, which are farther from the real concept meaning. However, our experiments have shown that using up to two hypernyms levels gets reasonably results. Using more levels decreases the quality, since obtained synset meaning is too far from the original intended meaning.

function selectCorrectSenseUsingMultilingualLabels.
input : concept *//concept to align*
output : senses *//selected senses*

for *label ∈ getLabels(concept)* **do**
   senses, processLevel ← getWordNetSense(label);
   sensesPerLabel.add(label, getLang(label),senses,processLevel);
end
processLevel ← minProcessLabel(sensesPerLabel);
sharedSenses ← ∅;
numHypernyms ← 0;
while *numHypernyms ∈ 0..2 & sharedSenses=∅* **do**
   bestSensesPerLabel ← removeMoreProcessedSenses(sensesPerLabel,
    processLevel);
   sensesPerLang ← groupLabelsPerLanguage(sensesPerLabel);
   numLang ← numLanguages(concept);
   while *numLang ∈ numLanguages(concept)..2 & sharedSenses=∅* **do**
      sharedsenses ← intersectXofYLanguages(sensesPerLang,
       bestSensesPerLabel, num);
      numLang ← numLang-1;
   end
   if *sharedSenses=∅* **then**
      sensesPerLabel.add(getWordNetHypernyms(sensesPerLabel));
      numHypernyms ← numHypernyms+1;
   end
end
return sharedSenses;

    Algorithm 2: Select correct sense using multilingual labels

If the intersection is multiple, the obtained synsets are additionally processed to identify the correct one. This is done using as disambiguation context the concepts previously aligned (see Algorithm 3). Since the terminology in a thesaurus is expected to be closely related, from the different possible senses we select the one that is closest in terms of geodesic distance

function selectCorrectSenseUsingContext.
input : candidateSenses *//possible senses of a concept - WordNet synsets*
              usedSenses *//senses used as alignment in other concepts*
output : refinedSenses *//reduced set of senses - ideally one sense*

for *sense ∈ candidateSenses* **do**
    closestSenses ← getSensesAtDistance(sense,3);
    closestSenses ← usedSenses ∩ closestSenses;
    weight ← 0;
    for *cSense ∈ closestSenses* **do**
        if *getDistance(sense, cSense)<=3* **then**
            weight ← weight+1/getDistance(sense, cSense);
        end
    end
    synsetDist.add(synset, weight);
end
refinedSynsets ← getMaximumWeightSynsets(synsetDist);
return refinedSynsets;

Algorithm 3: Select correct sense using context

(the length of any shortest path between two vertices) to the already aligned synsets, considering the is-a hierarchy of synsets as a graph. We consider aligned concepts up to a maximum distance of 3 of each synset (this includes brothers -dist=2- and uncles -dist=3-). As the weight of a synset, we use the sum of the inverse of the distances to already used synsets. The one with the higher weight is selected. For example, figure 3 shows the disambiguation of "street" between "a thoroughfare lined with buildings" and "depressed environment of poverty", supposing that "road" and "main road" synsets have been selected for other concepts in the thesaurus. In this case, it can be observed that the distances between "road", "main road" and the first "street" meaning is two and three respectively in the first case (weight =1/2+1/3), and greater than 3 in the second one (weight =0+0). We add a limit of distance since we think that too far concepts are not relevant to distinguish between small sense differences and distort the obtained measure. In our experiments, the use of a maximum distance of 3 provides the best results. In addition, it may happen that two or more possible synsets have the same weight. In this case, the concept is left unaligned.

Once a concept has been aligned with WordNet, the alignment with DOLCE is a simple task. We have updated an existent manual alignment between WordNet 1.6 and DOLCE [7] that matches the upper levels of WordNet
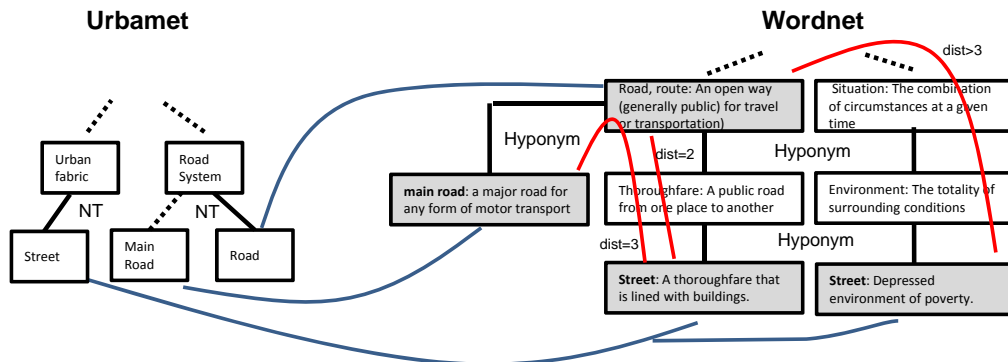
Figure 3: Example of context based disambiguation

as subclasses of DOLCE. Since hypernym/hyponym relations have a subclass meaning, all the hyponyms of each WordNet sense aligned with DOLCE (the upper ones) share the DOLCE meaning of their hypernyms and therefore, they are directly aligned to it. For example, in WordNet, the concept "street-light" is hyponym of "device". Since "device" is aligned to DOLCE "physical object", all of their hyponyms, including "streetlight", are also tagged as "physical object".

3.2.2. Verification of BT/NT relations

ISO 25964 states that BT/NT relations indicate superordination and subordination, where superordinate represents a class or whole and subordinate represents to its members or parts. Therefore, any BT/NT that cannot be interpreted according to this definition is incorrect. In WordNet, hypernyms (is-a) and meronyms (part-of) relations provide this superordination and subordination meaning. The same happens with some DOLCE relations such as "participant" or "exact location of". In this context, we use the existing relations in these two models to determine the correctness of the original relations in the thesaurus.

Figure 4 shows the four different types of relations we have analysed between concepts in a BT/NT relation and their corresponding synsets in WordNet. The examples use hypernyms relations, but the same is applicable to meronyms. We consider that two concepts aligned to WordNet with an exact match share the same relation that their corresponding WordNet synsets (1). The same happens if the subordinated concept is aligned with a broader match relation to WordNet (2). If both concepts are related to the same sense (3), the original relation is also considered correct. However,

14

nothing can be deduced when the superordinated concept is the one loosely matched (4), or if there is not a hierarchical connection between the concepts in WordNet.
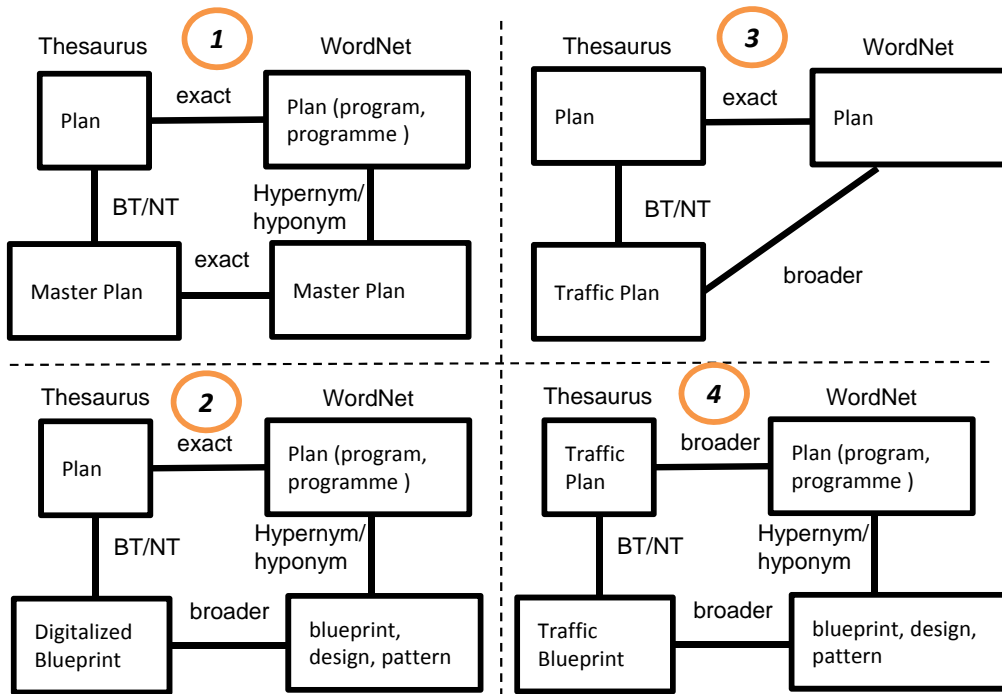

Figure 4: Possible relation refinement with WordNet

Since thesaurus terminology is usually quite specific and WordNet only covers general concepts, many of the superordinate concepts have "broader match" relations with WordNet (case 4 in figure 4). To validate these relations, we can use DOLCE ontology. DOLCE provides general relations between concept categories that we assume also hold between the concepts assigned to these categories. In this context, we have identified three families of DOLCE relations as compatible with BT/NT semantics: subclass, participation and location.

- The subclass relation indicates that the original concepts belong to hierarchically related categories. This does not ensure that the original BT/NT is correct, but because the thesaurus objective is to create generic to specific models, it is a good clue in that direction.

- The participation relation holds between perdurants (activities) and endurants (objects). It indicates elements that are part of an activity, which is a valid BT/NT meaning (e.g., horse piece is *part-of* a chess game or a car is *part-of* a car accident).

- With respect to the location related properties (from spatial to conceptual location), they may provide a *part-of* meaning (e.g., fountain *part-of* park) or an *is-a* meaning (e.g., linear town *is-a* kind of urban morphology), both valid in the thesaurus context. If a BT/NT relation is assigned to one of these families, we consider it as correct.

Figure 5 shows, as example, the validation of the relation between "Street" and "Crossing". Through the alignment with WordNet, they are identified as hyponyms of "Road" and "Path", which have been aligned to a "physical object" in DOLCE. Therefore, "Street" and "Crossing" are labelled as a kind of "physical object". In this case, most of the relations that may happen between two "physical objects" are subproperties of "material place of". Therefore, it is the one selected. Since this location-based relation provides a superordinate-subordinate meaning (a crossing is in a street, so it is part of the street), the original BT/NT is considered as valid.
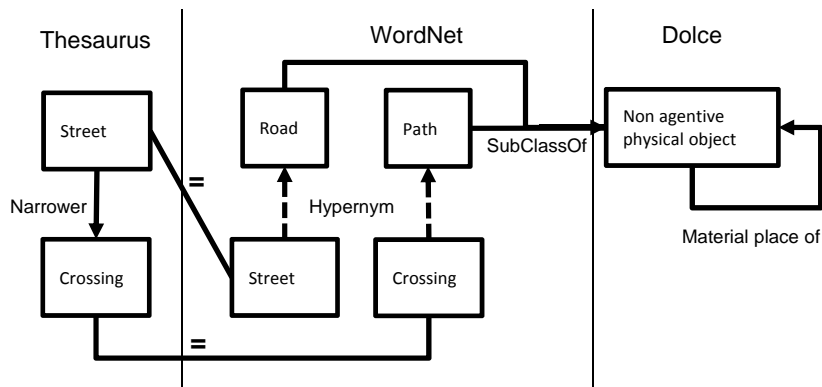

Figure 5: Example of validation of relations

4. Experiments

4.1. Description of the corpus

We have used the validation process to check the quality of two thesauri in different knowledge areas: Urbamet in the urbanism context, and Gemet in the environment field.

Urbamet is a thesaurus about urbanism created by the CDU (French Center of Urban Documentation); an organization focused on enhancing the exchange of urban knowledge and expertise between scientists, practioners and decision-makers. Designed and maintained by the CDU, the Urbamet database was initially developed in 1969 to facilitate the sharing of knowledge between professionals in urban planning, housing and transport in France. It gathers bibliographic notes issued internally in the center as well as those proposed by the 97 French departmental administrations in charge of equipment (roads, technical services and planning issues) and major urban planning agencies in France.

Gemet is a thesaurus about environment created by EIONET; a partnership network of the European Environment Agency, its member and cooperating countries. It consists of the EEA itself, six European Topic Centres and a network of around 1000 experts from 39 countries in over 350 national environment agencies and other bodies dealing with environmental information. GEMET was developed in 2001 as an indexing, retrieval and control tool for the European Topic Centre on Catalogue of Data Sources and the European Environment Agency.

We have analysed these two thesauri by processing their labels in English, French, and Spanish). Table 1 shows the main characteristics of the thesauri. It shows the number of concepts, preferred (PL) alternative labels (AL), and definitions (Def) in English, French, and Spanish, the number of broader/narrower relations, and related relations (RT). The columns with zero elements in both thesauri have been omitted.

Table 1: Details of Urbamet and Gemet thesauri

Thesaurus	Concepts	PL(en)	PL(es)	PL(fr)	AL(fr)	Def(en)	BT-NT	RT
Urbamet	3844	3844	3845	3846	504	0	3821	0
Gemet	5244	5244	5244	5244	0	4909	5332	1043

To be able to evaluate the quality detection process proposed in this, 5 reviewers have manually annotated the existent concepts and BT/NT relations in the "urban planning development" branch (208 concepts) of Urbamet, and the "biology" branch (310 concepts) of Gemet thesauri to determine the number of true/false positives and negatives. The only feature that has been problematic has been the *Semantic correctness of BT/NT* due to the subjectivity of some relations. For this feature, the judges have a proportion

agreement of 77.5% in Urbamet and 92.1% in Gemet. This gives an inter-agreement measure (Fleiss Kappa) of 0.42 (moderate agreement) in Urbamet and 0.36 in Gemet [16]. When there has been disagreement in a relation, we have selected as correct the one selected by most of the reviewers. The lower inter-agreement in Gemet is caused by the used measure. Fleiss Kappa penalizes each disagreement between the judges in a different way depending on the distribution of the collection. In Gemet, since there are very few incorrect relations, each disagreement between the judges is severely penalized. In the Urbamet case, the lack of agreement in so many relations can be also considered as an indicative of a problem in the structure.

4.2. Results of the quality analysis

The quality analysis performed by our tool has produced the results depicted in table 2. It shows the global degree of overall fulfilment of the measures described in the process description section in Urbamet and Gemet (percentage of correctness as explained in section 3).

Table 2: Report of Urbamet and Gemet thesaurus quality

Measure name	Urbamet value	Gemet value
Property completeness analysis		
Completeness and uniqueness of preferred labels	99.97%	100%
Completeness of definitions	0%	31.20%
Property content analysis		
Non-existence of non-alphabetic characters in labels	99.94%	99.92%
Nonuse of adverbs / initial articles in labels	**97.54%**	**98.45%**
Non use of acronyms in preferred labels	99.24%	99.28%
Property context analysis		
Non-existence of duplicated labels	99.77%	98.83%
Consistent use of uppercase in labels	99.35%	97.60%
Consistent use of plurals in labels	**94.50%**	**88.72%**
Property complexity analysis		
Nonuse of prepositional phrases in labels(en)	**90.01%**	**94.58%**
Nonuse of too long noun phrases / conjunctions	**99.30%**	**99.24%**
Relation coherence analysis		
Informative RTs	NA	93.67%
Completeness of BT/NT	99.40%	100%
Non-existence of BT/NT cycles	100%	100%
Semantic correctness of BT/NT	**69.70%**	**75.76%**

The results obtained indicate that both of the thesauri have issues in most of the analysed features (especially in the BT/NT structure). However,

some of these results cannot be completely trusted, as the processes used to detect them are not completely accurate (e.g., POS tagging is usually above 95% but not 100%). These features are highlighted with bold face in table 2.

Sections 4.3 and 4.4 review the results obtained for each measure and discuss about its accuracy. We distinguish between exact measures with perfect accuracy (existence checks or simple lexical comparisons), and approximate ones that include false positives and negatives.

4.3. Discussion on exact measures

In our context, the measures obtained without the use of stemming or POS tagging are exact and can be completely trusted. This includes the Completeness and uniqueness of preferred labels, Completeness of definitions, Non-existence of non-alphabetic characters in labels, Nonuse of acronyms in preferred labels, Consistent use of uppercase in labels, Non-existence of duplicated labels, Informative RTs, Completeness of BT/NT, and Non-existence of BT/NT cycles.

There are some important issues that have been detected in both thesauri thanks to this analysis. Focusing on Urbamet, we can first remark the lack of definitions, which makes difficult the correct use of the concepts for classification. It is also important that there are some incorrect labels with non-alphabetic characters. For example, the "snowfield" concept in Urbamet has as English label "snow DE_ENG", which seems to be an error issue related to the thesaurus creation. There are also numerous acronyms, and some uppercase labels. Additionally, there are some more important issues that should be observed such as the existence of duplicated concepts, the complete lack of RT relations and the existence of a few orphan concepts.

In Gemet the problems are similar. It contains definitions, but they are only available in English. There are also labels with non-alphabetic characters, acronyms, uppercase labels and duplicated concepts. Gemet has RT relations and a few of them are not informative (e.g., glue and adhesive are related with an RT although they belong to the same hierarchy). It does not have orphan concepts, but it contains concepts that do not have labels in all the languages.

4.4. Discussion on approximate measures

The measures highlighted in bold face in table 2 involve techniques that may introduce errors in the results. The detection of adverbs, nouns, con-

junctions, and prepositional phrases depends on the performance of the POS tagger to identify articles and adverbs. Something similar happens when detecting plurals, since the pattern based software used does not take into account all possible exceptions to general rules in order to distinguish plurals from singular. Finally, the detection of the correctness of BT/NT relations requires all the previous techniques, including additional heuristics for alignment and formalization.

In this section, we analyze the accuracy of these results in terms of false positives and negatives. Table 3 shows the result of this analysis. It includes the percentage of true/false positives (tp, fp) and negatives (tn, fn), and also the percentage of NT/BT relations that have not been reviewed because one of the concepts involved has not been aligned to WordNet (na). The accuracy of each measure has been calculated as the sum of true positives and negatives (tp+tn) divided by the number of elements in the collection (tp+tn+fp+fn+na).

Table 3: Performance of quality measure processes

Urbamet quality measure	tp	tn	fp	fn	na	accuracy
Nonuse of adverbs / initial articles in labels	99.67%	0%	0%	0.23%		99.67%
Consistent use of plurals in labels	97.75%	2.08%	0%	0.16%		99.83%
Nonuse of prepositional phrases in labels(en)	87.01%	12.98%	0%	0%		100%
Nonuse of too long noun phrases / conjunctions	99.16%	0.50%	0.33%	0%		99.66%
Semantic correctness of BT/NT	64.73%	12.56%	4.83%	14.49%	3.38%	77.29%

Gemet quality measure	tp	tn	fp	fn	na	accuracy
Nonuse of adverbs / initial articles in labels	99.73%	0.24%	0%	0.02%		99.98%
Consistent use of plurals in labels	92.91%	6.23%	0%	0.86%		99.14%
Nonuse of prepositional phrases in labels(en)	94.58%	5.42%	0%	0%		100%
Nonuse of too long noun phrases / conjunctions	99.90%	0.10%	0%	0%		100%
Semantic correctness of BT/NT	88.39%	1.30%	9.67%	0.64%	0%	89.69%

Except in the last case, the percentage of false positives and negatives is zero or almost zero. This was expected because we have used well known software libraries and algorithms that are known to have a good performance for natural language related tasks. However, some results are a bit misleading and require some additional comments.

The adverb and initial articles recognition works quite well for both thesauri, it identifies some errors, being the few false negatives single nouns

that the POS tagger classifies incorrectly as adverbs (due to the lack of context). The same happens with the detection of too long noun phrases or phrases with conjunctions, where the false positives in Urbamet correspond with words that the POS tagger incorrectly identifies as prepositions. The detection of plurals is also able to find most of the inconsistencies in its use, corresponding the few false positives with plural exceptions in English, Spanish and French. With respect to the prepositional phases in English, the ones that are identified correspond effectively with problematic concepts. Most of them can be replaced with a combination of adjectives and a substantive. For example, "master plan for water management" or "contractor working for the local council" can be replaced with "water management master plan" and "local council contractor". The rest describe concepts too complex and too specific to be applicable for classification. For example, the concept "code for tribunal dealing with internal disputes in the civil service" is a too long and complex concept, and it should be revised. Finally, related to the Semantic correctness of BT/NT, the process has tagged a relevant percentage of the BT/NTs relations as incorrect. This remarks the inherent difficulty of generating hierarchies according to thesauri standards. An example of incorrect relation is: "urban area" NT "master plan for urban development and planning zone" (a plan is neither an area nor part of an area nor an instance of an area). The quality of the results differs between the two analysed thesauri. In Urbamet, there are many false negatives. This indicates that correct relations are identified as erroneous. The problem in Gemet is the opposite, since many erroneous relations are not correctly identified. This can be caused by the differences in the terminology used in each thesaurus. In Urbamet, most of the concepts are compound nouns with a very general base noun (e.g., "master plan"). However, in Gemet they are simple nouns but very specialized (e.g., "enterovirus").

Since the analysis of the Semantic correctness of BT/NT relations is the one of the main contributions of the paper, we have performed a more detailed analysis of this feature. Specifically, we have analysed how the use of the multilingual labels of the concepts for concept disambiguation affects to the results. This is done by comparing the results of the multilingual process with the results obtained of applying the same process to monolingual versions of Urbamet and Gemet thesauri (only English, only French, and only Spanish). Table 4 shows the results obtained in each experiment. It includes the percentage of relations correctly identified (tp +tn), the incorrectly identified (fp +fn), and those not identified due to the lack of alignment of the

involved concepts.

Table 4: Semantic correctness of BT/NT

Urbamet Experiment	tp + tn	fp + fn	not analysed
Multilingual	77.29%	19.33%	3.38 %
English	56.04%	35.74%	8.22 %
Spanish	51.21%	21.73%	27.06%
French	61.35%	19.32%	19.33%
Gemet Experiment	tp + tn	fp + fn	not analysed
Multilingual	89.69%	10.31 %	0 %
English	88.70%	10.65%	0.65%
Spanish	15.49%	6.45 %	78.06%
French	26.13%	9.03 %	64.84%

These results show that the use of multilingual based concept disambiguation improves the performance of the process. In Gemet, there is only a small improvement with respect to the English version, but in Urbamet the improvement is clear. The problem with Gemet is related to the lack of completeness of Spanish and French versions of Wordnet, since they only contain a subset of the concepts contained in the English version.

Table 4 shows that, even in the best case, there is a relevant percentage of relations incorrectly identified. This is mainly related to errors in the alignment between the thesaurus and WordNet, because the incorrect alignment of a concept with WordNet usually leads to an incorrect alignment with Dolce (except if the erroneous sense selected is very similar to the correct one), and therefore to an incorrect analysis of the relations that involve the concept. Since inner concepts of a thesaurus participate in multiple relations, they affect to the identification of all the relations that involve it. For example, in the Urbamet experiment, we have manually verified that only 13.6% of the concepts have been incorrectly aligned with WordNet, but they participate in 24.5% of the relations. This highlights the importance of having a high quality thesaurus-WordNet alignment to be able to detect issues in the BT/NT relations with this process. It is also important to note that a few relations with concepts incorrectly aligned in WordNet end being correctly classified by the algorithm. This happens when the selected meaning is close to the correct one, so our process ends obtaining the same conclusion as if it were correct.

5. Related work

There are few systems that automatically analyse thesauri to determine their quality. Some of them focus on detecting structural issues, such as omitted language tags, or orphan concepts. In this context, Mader et al. [17] analyses the completeness of the properties and the syntactic coherence of the relations in a small collection of thesauri. Additionally, Suominen and Hyvönen [18] and Suominen and Mader [19] describe a set of tools (qSKOS, Skosify and PoolParty) that analyse a set of features similar to those proposed by Mader et al., but which also include automatic correction for some of types of the issues detected.

Other works focus on facilitating the detection of semantic issues in the concepts and relations. On the one hand, Spero [20] experiments with different visualizations of the Library of Congress Subject Headings broader terms to show how they can help users to judge which relationships might not be correct. On the other hand, Eckert [21] proposes an analysis of the suitability of the thesaurus concepts based on the use of the concepts in the associated collections. It analyses statistically the use of the thesaurus concepts in a collection and visualizes the usage results using a treemap. Although the focus of this work is to determine the thematic suitability of concepts for a given collection, it can be indirectly used for detecting the thesaurus quality. Irrelevant labels or those incorrectly located in the thesaurus hierarchy are expected to have few uses when classifying a collection. Additionally, Nohama et al. [22] identify anomalies in the concepts included in a thesaurus hierarchy using the idea that, in closely related corpora, the statistical distribution of semantic identifiers exhibits a high degree of correspondence. This approach allows detecting outlier labels of concepts that do not fit with the surrounding ones by analysing a collection of documents describing the thesaurus concepts terminology. Finally, Poveda-Villalón et al. [23] analyses ontology features according to six quality dimensions: human understanding, logical consistency, modelling issues, ontology language specification, real word representation and semantic applications. Although this tool is focused on ontology analysis, some of the features processed are applicable to thesaurus quality analysis (e.g., the detection of completeness of definitions, cycles in transitive relations, and orphan concepts).

In addition to these works directly focused on detecting quality issues in thesauri, there are works focused on re-engineering and formalization of thesauri that can indirectly be used for quality analysis. In this area, Tudhope

et al. [24] describe how to specialize the associative relations of the Art and Architecture Thesaurus (AAT) into richer subtypes through the analysis of sample extracts of AAT Editorial Related Term Sheets and the AAT editorial rules. Khosravi and Vazifedoost [25] propose a process based on transformation rules that allows the formalization of the ASFA Persian thesaurus relations. Focusing on other knowledge models, Aleksovski et al. [26] match lists of terms using different disambiguation and heuristic techniques and upper or domain formal ontologies as context. van Damme et al. [27] uses folksonomies and other unstructured vocabularies to construct ontologies. They derive ontologies from folksonomies based on statistical analysis, the use of online lexical and semantic web resources, the application of ontology matching (and mapping) approaches, and the computer assisted revision of the results. In all these works, the process to redefine the relations implies the need to understand their meaning, and therefore, it can be used to mark elements that do not fit in the thesaurus.

Table 5 shows the features analysed in our approach and compares them with respect to those used in other systems for analysing quality (explicitly or implicitly). Since each approach analyses slightly different features, they have been assigned to the closest category of the ones addressed in this paper. Our process determines the thesaurus quality performing structural, lexical, syntactical and semantical checks that validate the labels describing the concepts, and the provided relations. Structural, and lexical checks have been already analysed in works such as Mader et al. [17] or Suominen and Mader [19], but we also analyse the content, context and complexity of the properties. Our main contribution is to be able to deal with the semantic anomalies in concepts and BT/NT relations. Only Eckert [21] provides a method to detect issues in the relations, but at the cost of using a collection classified with the thesaurus. Poveda-Villalón et al. [23] also analyses some of the features processed in this paper, but the work is focused on ontologies and most of the included elements are not relevant for thesauri.

6. Conclusions and future work

This paper has described a process to automatically analyse the quality of thesauri according to the ISO 25964 standard doing a lexical and syntactic analysis of the concept labels, and a structural and semantic analysis of the properties and relations.

Table 5: Comparison of features analysed between approaches

	Mader et al. [17]	Suominen and Mader [19]	Eckert [21]	Poveda-Villalón et al. [23]	Our proposal
Property completeness analysis					x
Completeness and uniqueness of preferred labels	x	x			x
Completeness of definitions	x	x		x	x
Property content analysis					x
Non-existence of non-alphabetic characters in labels		x			x
Nonuse of adverbs / initial articles in labels					x
Nonuse of acronyms in preferred labels					x
Property context analysis					x
Non-existence of duplicated labels					x
Consistent use of uppercase in labels					x
Consistent use of plurals in labels					x
Property complexity analysis					x
Nonuse of prepositional phrases in labels(en)					x
Nonuse of too long noun phrases / conjunctions					x
Relation coherence analysis					x
Informative RTs	x	x			x
Completeness of BT/NT	x	x		x	x
Non-existence of BT/NT cycles	x	x		x	x
Semantic correctness of BT/NT			x		x

To determine if the meaning of the BT/NT relations is compatible with definition in ISO 25964, we have aligned it to WordNet and DOLCE ontologies and compared the original relations with the provided in these models. In many contexts, the existence of multilingual information is an issue, as they are an additional element to take into account to make a system work. However, for quality analysis of relations, we have seen how the use of the multilingual labels (English, Spanish and French) of the thesaurus concepts can be used to help to identify the correct sense in OMWordNet (WordNet multilingual version).

The results show the capacity of the process to identify existent problems in the structural, lexical, and syntactic facets. With respect to the semantic one, the results are mixed. It detects many incorrect relations, but also leaves a relevant number of relations incorrectly identified. This is mainly caused by

the errors in the alignment between the thesaurus and WordNet, and further improvement in this area would help to improve the final results.

In general, with the obtained results, it can be said that the proposed process is helpful to find anomalies in the models. Even if it generates some false positives and negatives, its ability to find real inconsistencies helps in the process of creating and reviewing a model. Additionally, it is important to note that the existence of an anomaly does not necessarily imply a real need to correct it. In some cases, the issues detected may be voluntary. What is not valid in a general context, it can be suitable in a given domain. However, even in this case it is relevant to detect them, as even the creators may not be conscious of certain aspects of their model.

Future work will focus on two main lines. On the one hand, we want to extend the process to perform alignments with multiple ontologies, which are already related between them, such as SUMO or openCYC, or even with other less formal models such as Wikipedia. The additional context provided by these models may help to improve the thesaurus hierarchical relations meaning. On the other hand, in addition to identifying BT/NT relations as incorrect, we also want to be able to identify the correct position of the concept in the thesaurus, i.e. identify the more appropriate broader term for each concept. Additionally, the method could also suggest "general" and "whole-part" specializations of the BT/NT relations (if supported by the original thesaurus) according to the alignments with other ontologies that have a richer semantics in their hierarchical relationships.

7. Acknowledgements

References

[1] International Organization for Standardization, Thesauri and Interoperability with other Vocabularies, ISO 25694, International Organization for Standardization (ISO), 2011.

[2] W. B. Frakes, R. Baeza-Yates (Eds.), Information Retrieval: Data Structures & Algorithms, chap. Thesaurus construction, Addison Wesley, 161–218, 1992.

[3] D. H. Fischer, From thesauri towards ontologies?, in: Structures and relations in knowledge organization - 5th International ISKO Conference, Lille (France), 18–30, 1998.

[4] D. Soergel, Indexing Languages and Thesauri: Construction and Maintenance, Melville Pub. Company, 1974.

[5] C. Maillard, URBAMET: le thesaurus et les lexiques. D'Urbamet à Urbadisc, Quaderni del CNBA. Coordinamento Nazionale Biblioteche di Architettura 3 (1998) 133–139.

[6] M. A. Hearst, WordNet: An Electronic Lexical Database, chap. Automated Discovery of WordNet Relations, MIT Press, 131–152, 1998.

[7] A. Gangemi, N. Guarino, C. Masolo, A. Oltramari, Sweetening WORDNET with DOLCE, AI Magazine 24 (3) (2003) 13–24.

[8] International Organization for Standarization, Quality management and quality assurance, ISO 8402, International Organization for Standarization, 1994.

[9] J. Aitchison, D. Bawden, A. Gilchrist, Thesaurus Construction and Use: A Practical Manual, Routledge, 2000.

[10] D. Kless, S. Milton, Towards Quality Measures for Evaluating Thesauri, Communications in Computer and Information Science. Metadata and Semantic Research 108 (2010) 312–319.

[11] M. Pinto, A user view of the factors affecting quality of thesauri in social science databases, Library & Information Science Research 30 (3) (2008) 216–221.

[12] C. Mader, B. Haslhofer, Perception and relevance of quality issues in web vocabularies, in: I-SEMANTICS'13 Proceedings of the 9th International Conference on Semantic Systems, 2013.

[13] J. Savoy, Report on CLEF-2001 Experiments, Tech. Rep., Institut interfacultaire d'informatique, Université de Neuchtel, Switzerland, 2001.

[14] R. E. Tarjan, Depth-first search and linear graph algorithms, SIAM Journal on Computing 1 (2) (1972) 146–160.

[15] J. Lacasta, J. Nogueras-Iso, G. Falquet, J. Teller, F. Zarazaga-Soria, Design and evaluation of a semantic enrichment process for bibliographic databases, Data & Knowledge Engineering 88 (2013) 94–107.

[16] J. L. Fleiss, Measuring nominal scale agreement among many raters, Psychological Bulletin 76 (5) (1971) 378–382.

[17] C. Mader, B. Haslhofer, A. Isaac, Finding Quality Issues in SKOS Vocabularies, in: P. Zaphiris, G. Buchanan, E. Rasmussen, F. Loizides (Eds.), Theory and Practice of Digital Libraries, vol. 7489 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 222–233, 2012.

[18] O. Suominen, E. Hyvönen, Improving the Quality of SKOS Vocabularies with Skosify, in: Knowledge Engineering and Knowledge Management, vol. 7603 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 383–397, 2012.

[19] O. Suominen, C. Mader, Assessing and Improving the Quality of SKOS Vocabularies, Journal on Data Semantics 3 (1) (2014) 47–73.

[20] S. Spero, LCSH is to Thesaurus as Doorbell is to Mammal: Visualizing Structural Problems in the Library of Congress Subject Headings., in: Proc. Int. Conf. on Dublin Core and Metadata Applications (DC), 2008.

[21] K. Eckert, Usage-driven Maintenance of Knowledge Organization Systems, Ph.D. thesis, Universitat Mannheim, 2012.

[22] P. Nohama, E. J. Pacheco, R. L. de Andrade, J. L. Bitencourt, K. Markó, S. Schulz, Quality issues in thesaurus building: A case study from the medical domain, Revista brasileira de engenharia biomedica .

[23] M. Poveda-Villalón, M. C. Suárez-Figueroa, A. Gómez-Pérez, Validating Ontologies with OOPS!, in: 18th International Conference EKAW, vol. 7603 of *Lecture Notes in Computer Science*, Springer, 267–281, 2012.

[24] D. Tudhope, H. Alani, C. Jones, Augmenting Thesaurus Relationships: Possibilities for Retrieval, Journal of Digital Information 1 (8) (2001) 15.

[25] F. Khosravi, A. Vazifedoost, Creating a Persian ontology through thesaurus reengineering for organizing the Digital Library of the National

Library of Iran, in: International Conference on Libraries, Information and Society, ICoLIS 2007, Petaling Jaya (Malaysia), 19–36, 2007.

[26] Z. Aleksovski, M. Klein, W. ten Kate, F. van Harmelen, Matching Unstructured Vocabularies using a Background Ontology, Lecture Notes in Computer Science 4248 (2006) 182–197.

[27] C. van Damme, M. Hepp, K. Siorpaes, FolksOntology: An Integrated Approach for Turning Folksonomies into Ontologies, in: Bridging the Gap between Semantic Web and Web 2.0 Workshop. ESWC 2007, Innsbruck (Austria), 57–70, 2007.