A Taxonomic Generalization Technique for Natural Language Processing

S. Ferilli[1,2], N. Di Mauro[1,2], T.M.A. Basile[1], and F. Esposito[1,2]

[1] Dipartimento di Informatica – Università di Bari
{ferilli, ndm, basile, esposito}@di.uniba.it
[2] Centro Interdipartimentale per la Logica e sue Applicazioni – Università di Bari

Abstract. Automatic processing of text documents requires techniques that can go beyond the lexical level, and are able to handle the semantics underlying natural language sentences. A support for such techniques can be provided by taxonomies that connect terms to the underlying concepts, and concepts to each other according to different kinds of relationships. An outstanding example of such a kind of resources is WordNet. On the other hand, whenever automatic inferences are to be made on a given domain, a generalization technique, and corresponding operational procedures, are needed. This paper proposes a generalization technique for taxonomic information and applies it to WordNet, providing examples that prove its behavior to be sensible and effective.

1 Introduction

Due to the complexity of natural language, most NLP techniques in the literature have so far focused on lexical-level representations such as bags of words. Unfortunately, using a strictly syntactical approach to text processing is often insufficient, because the words make direct and explicit reference to underlying concepts that have complex interactions and relationships to each other, and that are fundamental to understand the text and to relate texts to each other. For instance, two sentences such as “the man bought a car” and “the woman won a bicycle” would be considered as having nothing in common, while any human would immediately grasp their generalization of “a person acquiring a means of transportation”. Hence, the need to introduce and exploit taxonomic knowledge, as provided by existing lexical resources and ontologies. In this work, we will focus on the exploitation of taxonomic resources to set up a generalization framework. This involves two components: a procedure that, given two (sets of) words or concepts returns their generalization, and a procedure that, given a taxonomic model (possibly coming from previous generalizations of words/concepts) and an observed word/concept, checks whether the former covers the latter. In this paper, we will use a widely-known general-purpose lexical taxonomy, WordNet, as a sample resource on which demonstrating the proposed technique. However, it should be noted that this decision is just driven by the opportunity of having a wide-scope, readily-available taxonomic resource for testing the technique. Indeed, the technique itself is completely general and can be applied to any other (possibly domain-specific) resource having very general features provided by WordNet: the generalization-specialization relationship (for nouns and verbs) and some kind of relationship expressing closeness or similarity among taxonomic items (such as synonymy).

After providing some background and fundamental notions in the next section, Section 3 describes the approach to generalization and coverage, while Section 4 provides a qualitative validation of the approach. Finally, Section 5 concludes the paper and proposes future work issues.

2 Background

This section discusses the basic features of an ontology that are needed to apply the generalization technique proposed in the following sections. Although these features are general, they will be framed in the WordNet environment to have a more practical example. WordNet [1, 2] is a famous lexical taxonomy/ontology[1] inspired by psycho-linguistic theories on human memory. It takes into account two main concepts: *word forms*, i.e. their written aspect, and *word meanings*, i.e. their underlying concept. Differently from classical dictionaries, terms in WordNet are not organized as an alphabetically ordered list, but arranged in a graph determined by various kinds of relationships. Nodes representing terms are linked to the nodes representing the corresponding meanings. From the opposite perspective, each node representing a meaning is connected to all synonymous words expressing that meaning, this way defining the fundamental concept of *synset* ('synonymous set'). Synsets can be considered as unique identifiers for meanings (in the rest of this paper, the terms 'concept', 'meaning', 'sense' and 'synset' will be used interchangeably), and a textual definition, called a *gloss*, is also provided for each of them. Clearly, due to polysemy one term might have different meanings (i.e., be associated to many synsets), and might even belong to different syntactic categories. The current version of WordNet (3.0) includes more than 150.000 lexical forms and approximately 120.000 synsets.

Two kinds of relationships are defined in WordNet. Semantic ones always relate two synsets/meanings, while lexical relationships, on the other hand, involve both terms and synsets [3]. For the purposes of our technique, we need to focus just on the following semantic relationships:

Hyperonymy determines a generalization hierarchy on synsets, and is defined on nouns and verbs. It links a synset X to a more general one Y such that "X is a kind of Y". Interpreting it the other way round, one obtains its opposite relationship, *hyponymy*, that links a concept Y to a more specialized one X and hence determines specialization hierarchies. They are the largest relationships in WordNet.

Similarity used among adjectives only, according to the relationship of antinomy. The main adjectives on which such a relationship is set are called *head*

[1] There is a debate about the latter definition, since some require an ontology to contain formal definitions of the concepts.

synsets, and in turn are connected to similar *satellite synsets* that somehow inherit the antinomy relationships from the main meaning to which they are connected.

and on the following lexical ones:

Sinonymy is, as already pointed out, the main relationship in WordNet. Interestingly, synonymous terms are not directly connected to each other, but they are connected to the synset representing the underlying meaning. Thus, two terms are implicitly synonyms if both are linked to the same synset. Among several possible definitions of synonymy, WordNet adopts a perspective according to which two terms are synonyms if they can be safely replaced to each other in a given linguistic context without changing the sentence meaning. This clearly avoids the possibility of two terms in different syntactic categories being synonymous, and in practice neatly divides the whole taxonomy into four separate parts, corresponding to the syntactic categories of nouns, verbs, adjectives and adverbs.

See also a lexical relationship connecting related terms such that the latter helps in defining or understanding the former.

Of course, availability of additional relationships in the taxonomy, although not required, can allow to extend and refine the proposed technique if suitably exploited.

WordNet has gained a lot of attention in the literature, as a wide-coverage, general-purpose linguistic resource that tries to bridge the gap between the lexical level and the underlying semantics. Several translations (both in different languages and cross-language), tailorings to specific application domains, and extensions with additional information (e.g., WordNet Domains [4]) have been carried out. It has been thoroughly exploited for many tasks, among which Word Sense Disambiguation [5] and similarity assessment among concepts [6, 7, 8]. However, not much work seems to be available concerning inference strategies defined on the WordNet taxonomy to exploit it as a support for reasoning about (terms,) concepts and their relationships.

3 Taxonomic Generalization and Coverage

The problem we will face in this work can be defined as follows: given two concepts in a WordNet-like taxonomy, how to define their generalization. For instance, such a generalization may act as a model to be checked against further observed concepts. Hence, the problem of how to assess whether a model accounts for an observation. This setting can be extended taking into account sets of concepts instead of single concepts. This allows to handle words, even without any hint about their grammatical role and exact meaning, by replacing them with the set of their possible associated concepts in any grammatical category. In WordNet, nouns/verbs on one hand, and adjectives/adverbs on the other, have a very similar organization and relationships involving them. Accordingly, we will devise two generalization/coverage strategies.

For nouns and verbs a hierarchical generalization based on the Hyponymy relationship is implicitly available. The ancestors of a concept according to such a relationship can be interpreted as all its possible generalizations. A classical approach has been to take as a generalization of given concepts their Least Common Subsumer (i.e., the closest common ancestor). This might be misleading when the taxonomy is actually a heterarchy, where there might be several incompatible Least Common Subsumers. An alternative naive approach might consist in taking as a generalization of two synsets the set of all their common ancestors, and in saying that a model of this kind covers an observation term/synset if and only if there is a non-empty intersection between the model and the set of ancestors of the observation. However, this would be very loose, because there is a highest chance that the top (i.e., the most general and abstract) concepts in the hierarchy appear in both, resulting in the coverage of almost everything. Indeed, in such a hierarchy the closest concept that generalizes two given concepts is almost always too general to be of use (often just the root of the hierarchy, 'Entity') and might not be unique (due to polysemy). The problem is that the top-level ancestors are very general, and hence useless in practice because they would be over-generalizations. A solution might be ignoring (i.e., removing from the hierarchy) the very top concept 'Entity', or even the highest levels in the hierarchy, but this would introduce the problem of how to determine up to which level to ignore, and how being sure that the removed levels are in fact irrelevant to the task correctness. To avoid this problem, a baseline approach that we will adopt consists in assuming that a model covers an observation if and only if the ancestors of the observation include all of the model synsets. It is clearly a very cautious/pessimistic approach (requiring that the set of synsets in the model is a subset of the set of ancestors of the observation is a very strict bias), but can serve as a reference for comparing the performance of other solutions.

We propose a generalization technique that selects, among all common ancestors of the two elements to be generalized, the 'border' set of all 'minimal' ancestors (in the sense that they have no descendant in the set of common ancestors, a kind of leaves of such a hierarchy). Given a set of concepts X, let us denote by B_X the border of X, and by A_X the remaining ancestors of X. In some sense, the border represents the set of all minimally general generalizations, resembling for this the version space approach [9]. Considering only the 'border' subset, the model is not more general than needed and hence does not include concepts more general than those it should account for. In this way, the initial option of checking for a non-empty intersection between the ancestors of the observation (including the synsets in the observation itself) and the border synsets in the model becomes much more sensible. The underlying rationale is that the model specificity is expressed by its border synsets, and not by all of its ancestors, and hence only the former should be exploited for checking observation coverage. More formally: given a model M and an observation O, M *covers* O if $B_M \cap (B_O \cup A_O) \neq \emptyset$, i.e. at least one of the meanings in the model covers (i.e., is in the generalization hierarchy of) one of the synsets in the observation. Using B_M (instead of $B_M \cup A_M$, as in the naive version) avoids that

Algorithm 1 Generalization technique for nouns and verbs in a taxonomy

```
Require: X: set of concepts
Require: T: taxonomy
  $A \leftarrow$ concepts($T$) /* set of common ancestors */
  for all $x \in X$ do
    $A \leftarrow A \cap \{a \in \text{concepts}(T) \mid \text{ancestor}_T(a, x)\}$
  end for
  return $\{c \in A \mid \nexists c' \in A \text{ s.t. } \text{ancestor}_T(c', c)\}$
```

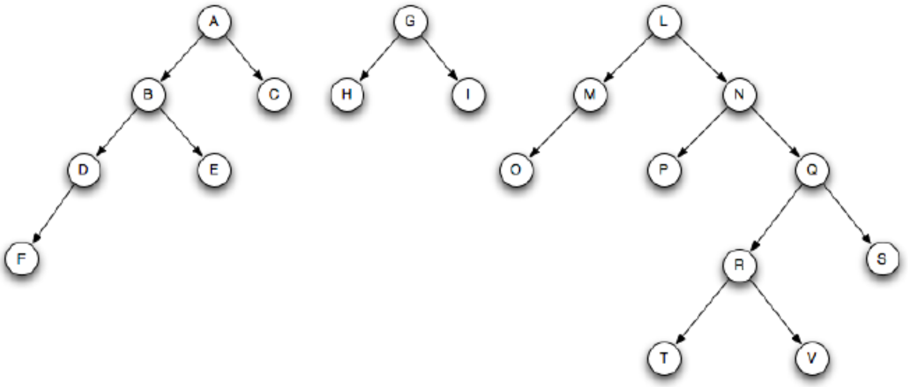

Fig. 1. A hypothetical fragment of taxonomy.

the observation is covered by an ancestor in the model and hence, actually, by an over-generalization, and with respect to the baseline model is not so strict as to require that all of the model synsets are included in the ancestors of the observation. Algorithm 1 sketches the pseudo-code of the procedure, assuming a taxonomy on whose elements an 'ancestor' relationship is defined (corresponding to the transitive closure of the generalization relationship).

As an example, consider the tree in Figure 1, and two words P_1 and P_2, corresponding respectively to synsets (nodes) $\{F, H, T\}$ and $\{E, I, S\}$. Their generalization M is equal to $\{B, A, G, Q, N, L\}$, where the border of M is $B_M = \{B, G, Q\}$ and the set of ancestors of M is $A_M = \{A, N, L\}$. This is the chosen model. Now, let us consider word P_3, corresponding to the set of synsets $\{D, R\}$: the set $(B_{P_3} \cup A_{P_3})$ including both its nodes and the ancestors of its nodes is $\{D, B, A, R, Q, N, L\}$. Thus, since $(B_{P_3} \cup A_{P_3}) \cap B_M \neq \emptyset$, P_3 is covered by M. Finally, consider word P_4, corresponding to the set of nodes $\{C, P\}$. The set $(B_{P_4} \cup A_{P_4})$ comprising its nodes and their ancestors is $\{C, A, P, N, L\}$. Since $(B_{P_4} \cup A_{P_4}) \cap B_M = \emptyset$, P_4 is not covered by M.

As another example actually taken from WordNet, generalizing 'pencil' and 'rubber' (both can be interpreted both as tools and as the underlying matter), the baseline generalization would be:

[entity], [whole,whole_thing,unit], [object,physical_object], [substance,matter], [artifact,artefact], [implement], [instrumentality,instrumentation]

whose border is:

[substance,matter], [implement]

Adjectives are not hierarchically organized as nouns and verbs. E.g., although 'colored' can be considered as a generalization of 'red', such a relationship is not specified in WordNet. The following relationships are available in WordNet for adjectives: Similarity, Attribute, See also, Participial, Pertinence, Derivation. We propose to select the set of items connected by the 'Similarity' and 'See also' relationships to the two adjectives to be generalized, because in a sense they express all possible variations thereof, and then taking their intersection. Since in this case the generalization does not consist of more abstract concepts, but of closely related ones, the coverage algorithm for adjectives consists in checking that there is a non-empty intersection between the adjectives in the model and the set of 'ways for defining' the corresponding synset in the observation. I.e., at least an element in the model must be related by similarity (relationship 'Similarity') or must provide further information (through relationship 'See also') to the adjective synset in the observation. This strategy can be applied also to adverbs derived from adjectives, by switching to the corresponding adjectives, while no hint is available for the others.

4 Evaluation

To evaluate the viability of the proposed techniques, several groups of 6 words for each word category were chosen from WordNet3.0 by linguistic experts, to which a concept of term generalization was provided, but who were not aware of the specific algorithm discussed above. Specifically, for each group they selected two reference words to be generalized, and four more test words: two somehow related to the reference words (that, in principle, had to be covered by the generalization), and two conceptually unrelated (that should not be covered by the generalization). This setting was devised to provide indications of the generalization and coverage behavior on both false positives and false negatives, for evaluating both completeness and consistency. The performance of the generalization was then compared to the naive baseline. In the following, for each group of test words (along with an explanation for the choice of such words, when useful), a table will show the behavior of the proposed technique against the baseline (where Y means that the generalization covers the observation, while N means that it does not).

First, a few cases taken from the category of nouns were considered. In Case 1, there are two tricky tests: 'boy', that is to be covered with reference to an application of the generalized terms to persons, and 'antilope', that might be misleading being an animal, but must not be covered being a herbivore.

Case 1 : dog-cat

Term	theoretical	motivation	proposal	baseline
leopard	Y	a carnivore	Y	N
boy	Y	cat and dog are used also referred to persons	Y	N
antilope	N	a herbivore	N	N
book	N		N	N

Case 2 : nail-hammer

Term	theoretical	motivation	proposal	baseline
hand	Y	nail and hammer (a bone in the ear) are parts of the body	Y	N
saw	Y	nail and hammer are carpentry tools	Y	N
girl	N		N	N
river	N		N	N

Some examples from the category of verbs are reported below. Specifically, Case 2 is particularly interesting due to the highly polysemic behavior of 'play'.

Case 1 : introduce-repose

Term	theoretical	motivation	proposal	baseline
insert	Y	putting something in	Y	N
enclose	Y	putting something in a container	Y	N
change	N		N	N
increase	N		N	N

Case 2 : play-represent

Term	theoretical	motivation	proposal	baseline
pretend	Y	indicates an artificial behavior, like in drama acting	Y	N
perform	Y	to put on a show or performance	Y	N
swim	N		N	N
think	N		N	N

Then, for adjectives, words that in at least one of their meanings can be considered similar to the reference pair were taken into account as positive test cases. Interestingly, in Case 1 coverage is correctly recognized also for test cases that refer to different perspectives on the reference terms ('incisive' is an abstract interpretation, while 'needlelike' is more concrete).

Case 1 : sharp-acute

Term	theoretical	motivation	proposal	baseline
incisive	Y	referred to the effectiveness of a reasoning, way of thinking or of speaking	Y	Y
needlelike	Y	somehow in-between the two reference terms	Y	Y
happy	N	a state-of-mind	N	N
hungry	N	a psycho-physical state	N	N

Case 2 : tiny-little

Term	theoretical	motivation	proposal	baseline
young	Y	has a similar meaning to the reference terms when referred to persons	N	N
small	Y	has a similar meaning to the reference terms when referred to persons or things	Y	Y
depressed	N	a state-of-mind	N	N
long	N	referred to distance rather than size	N	N

Finally, some adverbs derived from adjectives were considered, using the same rationale as reported above for adjectives.

Case 1 : quickly-rapidly

Term	theoretical	motivation	proposal	baseline
speedily	Y	a synonym	Y	N
apace	Y	a synonym	Y	N
near	N		N	N
easily	N		N	N

Case 2 : appropriately-fittingly

Term	theoretical	motivation	proposal	baseline
suitably	Y	a synonym	Y	N
fitly	Y	a synonym	Y	N
jointly	N		N	N
playfully	N		N	N

Table 1 summarizes the results for the complete set of word groups, and for each word category, both overall and on positive/negative cases only. It clearly emerges that the proposed solution not only represents an outstanding improvement on the baseline, but produces very high-quality results in itself. More specifically, although the generalizations are more compact in the proposed procedure due to the focus on the border only, the key for the improvement is represented by the coverage procedure. Indeed, as to nouns and verbs, the baseline coverage strategy is very pessimistic, and rejects almost all test words: all negative cases are correctly rejected, but no positive cases are covered for nouns, nor for verbs. Conversely, the proposed technique provides a perfect behavior on both positive and negative cases for these categories. On adjectives and adverbs the generalization and coverage strategies are the same for both the 'baseline' and the 'proposal', but two different evaluation strategies were compared: in the former, pairwise comparisons are carried out among single meanings in the two references, while the latter adopts a global approach that first expands all meanings, and only subsequently intersects the resulting sets as a whole. Actually, on adverbs the proposed approach turns out to be (significantly) better. Conversely, as to adjectives, the coverage performance of the proposed approach is just the

Table 1. Statistics on performance (accuracy) of the taxonomic generalization and coverage procedures

Coverage	Proposal		Baseline		Improvement
All	31/32	97%	19/32	59%	38%
positive only	15/16	94%	3/16	19%	75%
negative only	16/16	100%	16/16	100%	0%
Nouns	8/8	100%	4/8	50%	50%
positive only	4/4	100%	0/4	0%	100%
negative only	4/4	100%	4/4	100%	0%
Verbs	8/8	100%	4/8	50%	50%
positive only	4/4	100%	0/4	0%	100%
negative only	4/4	100%	4/4	100%	0%
Adjectives	7/8	87%	7/8	87%	0%
positive only	3/4	75%	3/4	75%	0%
negative only	4/4	100%	4/4	100%	0%
Adverbs	8/8	100%	4/8	50%	50%
positive only	4/4	100%	0/4	0%	100%
negative only	4/4	100%	4/4	100%	0%

same as the baseline, although it should be said that it could be hardly improved because only one error on positive cases occurs.

5 Conclusions

Several techniques for processing texts are based on the lexical level in order to make the problem computationally tractable. However, the tricks of natural language, and the relationships among the concepts underlying the words that are used in text, call for some kind of taxonomic background knowledge to help handling the semantics underlying the sentences. An outstanding example of such a kind of resources is WordNet. This paper proposed a generalization procedure, and a coverage procedure, to be exploited for automatic inferences on a given domain for which basic taxonomic information is provided. Although the problem is not very suitable to statistic evaluation, selected test cases were devised, and the outcome of the proposed technique on WordNet terms revealed that its behavior is sensible and effective.

Future work, part of which is already ongoing, includes running wider and more varied experiments. Moreover, further relationships expressed in WordNet might be exploited for improving the generalization. The proposed technique might be exploited to extend a purely syntactical approach to inference in First-Order Logic, where some predicates are not just uninterpreted syntactic entities, but are associated to nodes in a taxonomy, which would allow to exploit the relationships in the taxonomy as a background knowledge in order to tackle more complex problems. In particular, we intend to embed it in an existing framework for symbolic Machine Learning presented in [10, 11]. A possible example of application might be learning from structural representations of natural language sentences, as proposed in [12].

References

[1] Miller, G.A., Beckwith, R., Fellbaum, C., Miller, K., Gross, D.: Introduction to wordnet: An on-line lexical database. International Journal of Lexicography **3**(4) (1990) 235–244

[2] Fellbaum, C.: WordNet: An Electronic Lexical Database. MIT Press, Cambridge, MA (1998)

[3] Miller, G.A.: Wordnet: a lexical database for english. Commun. ACM **38** (November 1995) 39–41

[4] Bentivogli, L., Forner, P., Magnini, B., Pianta, E.: Revising the wordnet domains hierarchy: semantics, coverage and balancing. In: Proceedings of the Workshop on Multilingual Linguistic Ressources. MLR '04, Stroudsburg, PA, USA, Association for Computational Linguistics (2004) 101–108

[5] Banerjee, S., Pedersen, T.: An adapted lesk algorithm for word sense disambiguation using wordnet. In: Proceedings of the Third International Conference on Computational Linguistics and Intelligent Text Processing. CICLing '02, London, UK, UK, Springer-Verlag (2002) 136–145

[6] Resnik, P.: Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural language. Journal of artificial intelligence research **11** (1999) 93–130

[7] Budanitsky, A., Hirst, G.: Semantic distance in wordnet: An experimental, application-oriented evaluation of five measures. In: Proc. Workshop on WordNet and Other Lexical Resources, 2nd meeting of the North American Chapter of the Association for Computational Linguistics, Pittsburgh (2001)

[8] Ferilli, S., Biba, M., Di Mauro, N., Basile, T.M., Esposito, F.: Plugging taxonomic similarity in first-order logic horn clauses comparison. In Serra, R., Cucchiara, R., eds.: AI*IA 2009: Emergent Perspectives in Artificial Intelligence. Volume 5883 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg (2009) 131–140

[9] Mitchell, T.M.: Version spaces: a candidate elimination approach to rule learning. In: Proceedings of the 5th international joint conference on Artificial intelligence - Volume 1, San Francisco, CA, USA, Morgan Kaufmann Publishers Inc. (1977) 305–310

[10] Semeraro, G., Esposito, F., Malerba, D., Fanizzi, N., Ferilli, S.: A logic framework for the incremental inductive synthesis of datalog theories. In Fuchs, N., ed.: Logic Program Synthesis and Transformation. Volume 1463 of Lecture Notes in Computer Science. Springer Berlin-Heidelberg (1998) 300–321

[11] Ferilli, S., Basile, T.M.A., Biba, M., Di Mauro, N., Esposito, F.: A general similarity framework for horn clause logic. Fundamenta Informaticae **90** (January 2009) 43–66

[12] Ferilli, S., Fanizzi, N., Semeraro, G.: Learning logic models for automated text categorization. In: Proceedings of the 7th Congress of the Italian Association for Artificial Intelligence on Advances in Artificial Intelligence. AI*IA 01, London, UK, Springer-Verlag (2001) 81–86