Enriching Ontologies by Learned Negation

or How to Teach Ontologies Vegetarianism

Daniel Fleischhacker

KR & KM Research Group, University of Mannheim, Germany
daniel@informatik.uni-mannheim.de

1 Problem Statement

Ontologies form the basis of the semantic web by providing knowledge on concepts, relations and instances. Unfortunately, the manual creation of ontologies is a time-intensive and hence expensive task. This leads to the so-called knowledge acquisition bottleneck being a major problem for a more widespread adoption of the semantic web. Ontology learning tries to widen the bottleneck by supporting human knowledge engineers in creating ontologies. For this purpose, knowledge is extracted from existing data sources and is transformed into ontologies. So far, most ontology learning approaches are limited to very basic types of ontologies consisting of concept hierarchies and relations but do not use large amounts of the expressivity ontologies provide.

Negation is of great importance in ontologies since many common ideas and concepts are only fully expressible using negation. An example for the usefulness of negation is the notion of a *vegetarian* who is characterized by *not* eating meat. It is impossible to fully formalize this notion without applying negation at some level. Not stating these additional information on vegetarians would severely limit the possibilities to deduce new knowledge on vegetarians from the ontology by doing reasoning. Furthermore, negation is of great significance for assessing the quality of ontologies. Without it, ontologies may never get incoherent or inconsistent which is an important quality criterion. Additionally, with negations contained in ontologies, it is possible to use ontology debugging approaches more effectively.

Given all these points, we consider it important to put effort into a more elaborate research of automatic or semi-automatic learning of negation for enriching ontologies.

2 State of the Art

There is a large number of possible data sources all of them exposing different properties with respect to their structure and content. To handle these different inputs, ***ontology learning*** makes use of approaches from many different research areas which leads to a wide spectrum of different ontology learning methods [2].

Regarding the learning of negation there is little work so far. An example being the ***extraction of concept disjointness*** as a special case of negation. Haase and Völker [6] use lexico-syntactic patterns and their work is extended by Völker et al. [15] applying classification approaches on a number of different lexical and structural features in their

LeDA tool. However, these approaches focus on the generation of disjointness of *atomic* classes and are not directly applicable for generating axioms containing complements of *complex* class expressions. Even if most negation axioms, i.e., axioms containing explicit negation, may be represented by disjointness, the representation in data (e.g., vegetarian as *someone who does not eat meat*) not necessarily resembles disjointness.

Another ontology learning method which also generates negation is implemented by the DL-Learner tool [10]. It uses ***inductive logic programming (ILP)*** to yield complex axioms describing concepts from a given ontology. Unfortunately, this method suffers from two issues. First, it is limited to using ontologies or data sets convertible to ontologies as data sources, thus it is not adequate to handle unstructured data and probably most semi-structured information. Secondly, the approach is not directly applicable to large data sets. This is mainly because of the dependency on reasoning for generating the relevant axioms which introduces scalability problems. Hellmann et al. [8] propose a method to extract fragments from the data to reduce it to a processable size but this could nevertheless lead to the loss of relevant data.

Texts are reasonable sources to extract knowledge about negation axioms, and detecting negation in texts could be a valid first step towards reaching this goal. Thus, work regarding the general ***detection of negation in biomedical texts*** is also of interest for learning negation. Most research in detecting negation in texts has been made in the biomedical domain where the approaches are used to extract data on the presence or absence of certain findings. This detection is mainly done by means of a list of negation markers and regular expressions [1], by additionally using linguistic approaches like grammatical parsing [5, 9] or by applying machine learning techniques [11–13]. It is particularly important that the detection of negation also requires the identification of its scope, i.e., the parts of the sentence the negation is referring to. Even if some of the mentioned works might be usable on open-domain texts, there is no evaluation in this direction but only for the biomedical domain and thus there is no information on their performance for other domains. Furthermore, it is not clear if detected negations are similar to the ones required in ontologies.

Recently, there has been more work on ***negation detection for open-domain texts*** mainly driven by its usefulness for sentiment analysis [3] or contradiction detection [7]. Councill et al., who particularly concentrate on the task of detecting the scopes of negation, also evaluated their approach on product review texts using an appropriate, annotated gold standard which unfortunately seems not to be publicly available. Despite these recent works, detecting negation in open-domain texts remains an open problem.

3 Expected Contributions

The main contribution of this work is expected to be the development of approaches to ***enrich given ontologies with negation axioms extracted from texts*** as a part of an overall ontology learning approach and accompanied by a corresponding implementation. For this purpose, we will take already existing, manually engineered ontologies and add negation axioms extracted from free texts.

Negations in ontologies could provide great benefit for many application. In the field of biomedicine, one example would be an ontology containing information on different drugs. For some of these drugs, it is known that there are bacteria which are resistant against them. For instance, *methicillin-resistant Staphylococcus aureus* (MRSA) are strains of *Staphylococcus aureus* resistant against beta-lactam antibiotics. To represent this in an ontology, the axiom $\texttt{BetaLactamAntibiotic} \sqsubseteq \neg\exists\texttt{effectiveAgainst.MRSA}$ could be used which is a complex negation. Given such negation axioms, it would be possible to deduce from the ontology which drugs are not suitable for treating diseases caused by specific pathogens.

A second contribution will be developing and employing approaches to ***combine multiple ways of extracting negation***. This will help compensating possible shortcoming of certain approaches or data sources and to achieve better overall results.

When enriching ontologies by negation, we have to pay special attention to the ***maintenance of the ontology's consistency and coherence***. Without this, there is the risk of rendering the ontology inconsistent and less useful for reasoning tasks. Such inconsistencies do not have to come from the addition of the learned negation axioms themselves but may also arise from erroneous non-negation axioms added by the overall learning approach.

To be able to actually evaluate the results gained by extracting negations from different data sources, an appropriate evaluation strategy is necessary. Based on related work, we will ***develop methodologies suited for the evaluation***.

4 Methodology and Approach

In the following, we give an overview on the methodology which we want to follow to come up with the aforementioned contributions.

Negation Extraction from Text We expect the detection of negation in textual data to be domain-dependent to a high degree. However, we will focus on the biomedical domain because of the large amount of work already done there regarding negation detection and the availability of expressive ontologies. There are several kinds of negations in texts which we will have to handle. Mostly, these kinds of textual negations are distinguishable into direct negation like caused by the word *not* and indirect negation recognizable by words like *doubt*, which introduce the negation solely by their semantics, and *misunderstanding*, where the semantics of negation is characterized by morphological markers like *mis-*. For the first manner of indirect negation, the lexical-semantic relation of antonymy may provide some additional hints for detection. This is why we already did experiments on detecting antonymy relations by means of relatedness and similarity measures. We will evaluate the approaches from the biomedical domain regarding their coverage for these different kinds of negation and develop approaches to treat the yet uncovered ones. To do this, we will most likely start with pattern-based detection approaches and then additionally apply machine learning methods.

For the enrichment of ontologies, we have to develop approaches to actually transfer the extracted textual negations into suitable logical negation which is not a trivial problem because of the ambiguity of natural language. Furthermore, we will evaluate the way negation is used in existing ontologies particularly regarding possible modeling errors made by humans and regarding the expressivity required for these negation axioms. Based on the findings, we will choose description logic fragments best suited for representing the learned negation while maintaining desirable computational properties.

An especially interesting approach is the combination of ways to learn from multiple data sources. As mentioned, this can help to compensate shortcomings in different approaches or data sources. LeDA [15] already combined different approaches but this is only done in course of their disjointness extraction algorithm and not directly applicable for combining arbitrary approaches. Having a more general way of combining different approaches, we could use it to integrate the negation axioms extracted by our proposed text-based system and other systems like DL-Learner [10].

Consistency Maintenance The task of consistency maintenance has to be employed for the overall ontology learning and enrichment process and not only for the enrichment by negation axioms. Most ontology learning approaches produce confidence values for generated axioms. Thus, we have to deal with uncertainty like Haase and Völker who also considered uncertainty to create consistent ontologies by ontology learning. We will apply similar debugging methods but also more general approaches like the one by Schlobach [14]. Regarding the overall learning approach, we will also explore methods of instantiating a feedback loop from debugging to the actual learning process. For the ontologies containing negation axioms, we are also able to compute different measures, e.g., the number of incoherent concepts widely seen as an indicator for an ontology's quality.

Evaluation The evaluation of the correctness of the created negation axioms is also important for the overall goal of learning negation. As there is no standard way of evaluating these axioms, we will propose a new methodology. There are different ways of evaluating general ontology learning approaches [4]. For negations, it seems to be less desirable to use a gold standard especially since its manual creation is extremely labor-intensive for large data sources. Alternatively, we could use the learning approach in an application which benefits from a more expressive ontology. For our evaluation, we will look for such applications. Finally, we could let human domain experts evaluate the extracted axioms regarding their correctness. Even if this means that there is no possibility of computing the completeness for the extracted axioms with respect to a given data source, important values such as the inter annotator agreement may still be computed.

5 Conclusion

In this paper, we presented our plans to develop and implement approaches to enrich ontologies by complex negation axioms. As described above, we consider this beneficial for a couple of reasons. Having the results in the area of negation detection for biomedical texts and some for open-domain texts, we already have some foundations regarding negations in texts which should enable us to achieve first results soon. All in all, learning approaches for negation can assist humans in creating more thoroughly formalized ontologies and thus lead to a more expressive semantic web.

References

1. Chapman, W.W., Bridewell, W., Hanbury, P., Cooper, G.F., Buchanan, B.G.: A simple algorithm for identifying negated findings and diseases in discharge summaries. Journal of Biomedical Informatics 34(5), 301 – 310 (2001)
2. Cimiano, P., Mädche, A., Staab, S., Völker, J.: Ontology learning. In: Staab, S., Studer, R. (eds.) Handbook on Ontologies, 2nd edition, pp. 245–267. Springer (2009)
3. Councill, I.G., McDonald, R., Velikovich, L.: What's great and what's not: learning to classify the scope of negation for improved sentiment analysis. In: Proc. of the Workshop on Negation and Speculation in Natural Language Processing. pp. 51–59 (2010)
4. Dellschaft, K., Staab, S.: On how to perform a gold standard based evaluation of ontology learning. In: Cruz, I. et al. (ed.) The Semantic Web - ISWC 2006, LNCS, vol. 4273, pp. 228–241. Springer (2006)
5. Gindl, S., Kaiser, K., Miksch, S.: Syntactical negation detection in clinical practice guidelines. Studies in Health Technology and Informatics 136, 187–192 (2008)
6. Haase, P., Völker, J.: Ontology learning and reasoning - dealing with uncertainty and inconsistency. In: Proc. of the Workshop on Uncertainty Reasoning for the Semantic Web (URSW). pp. 45–55 (2005)
7. Harabagiu, S., Hickl, A., Lacatusu, F.: Negation, contrast and contradiction in text processing. In: Proc. of the 21st national conference on Artificial intelligence - Volume 1. pp. 755–762 (2006)
8. Hellmann, S., Lehmann, J., Auer, S.: Learning of OWL class descriptions on very large knowledge bases. International Journal On Semantic Web and Information Systems 5, 25–48 (2009)
9. Huang, Y., Lowe, H.J.: A novel hybrid approach to automated negation detection in clinical radiology reports. Journal of the American Medical Informatics Association 14(3), 304–311 (2007)
10. Lehmann, J.: DL-Learner: Learning concepts in description logics. Journal of Machine Learning Research 10, 2639–2642 (2009)
11. Li, J., Zhou, G., Wang, H., Zhu, Q.: Learning the scope of negation via shallow semantic parsing. In: Proc. of the 23rd International Conference on Computational Linguistics. pp. 671–679 (2010)
12. Morante, R., Daelemans, W.: A metalearning approach to processing the scope of negation. In: Proc. of the 13th Conference on Computational Natural Language Learning. pp. 21–29 (2009)
13. Sarafraz, F., Nenadic, G.: Using SVMs with the command relation features to identify negated events in biomedical literature. In: Proc. of the Workshop on Negation and Speculation in Natural Language Processing. pp. 78–85 (2010)
14. Schlobach, S.: Debugging and semantic clarification by pinpointing. In: Gómez-Pérez, A., Euzenat, J. (eds.) The Semantic Web: Research and Applications, LNCS, vol. 3532, pp. 226–240. Springer (2005)
15. Völker, J., Vrandečić, D., Sure, Y., Hotho, A.: Learning disjointness. In: Franconi, E., Kifer, M., May, W. (eds.) The Semantic Web: Research and Applications, LNCS, vol. 4519, pp. 175–189. Springer (2007)