Chapter 6
Conclusions and Outlook

After laying the theoretical foundations in the first three chapters, and then presenting and evaluating the relation label suggestion methods in Chapters 4 and 5, this chapter concludes the thesis summarizing the approach and experiments conducted. It highlights the main contributions, and outlines the most promising areas for future research.

Summary. In ontology learning, the task of labeling non-taxonomic relations in domain ontology is among the most difficult and least tackled problems [95]. The presented approach introduces a set of methods to address this issue. This approach combines corpus-based methods, which have domain text as their only source of input, with a technique to validate ontological restrictions relying on knowledge inferred from Semantic Web information sources. The corpus-based methods utilize verbs co-occurring with the respective relations in vector space models to calculate the similarity to known relations. Based on the similarity values, the algorithms refine the relation labeling results by validating the conformance of the entities involved against ontological restrictions defined with the help of a meta ontology. The crucial ingredient in this process is concept grounding, i.e. the task of linking the concepts from the domain ontology into the meta ontology in a procedure that includes reasoning techniques with external data sources, such as DBpedia and OpenCyc.

An extensive set of experiments helped to assess the performance of the presented approach. Training and testing relations were labeled with one of five basic predicates. When distinguishing the correct predicate and the direction of the relation, this resulted in ten relation label candidates. The evaluation metrics of Average Ranking Precision, first guess correct and second guess correct, were applied to evaluate different configurations of the relation labeling method. The method yields an accuracy of 53% correct
suggestions on first guesses regarding relation type and direction. When ignoring direction the accuracy increases to 75%. The average position of the correct label in the list of label suggestions is about 2.0 with ten candidates, and slightly below 1.5 when neglecting relation direction.

The evaluation results fluctuate depending on the configuration used by the architecture. Some of the settings had no consistent positive or negative effect on performance, for example the inclusion of the 150 most significant verbs per relation in the verb vectors versus the 20 most significant, or confirming verbs with WordNet – the outcome depends on the remaining evaluation metrics and configurations chosen (as outlined in Chapter 5). Other settings, like the optional use of prepositions occurring directly after verbs in text, consistently yielded positive effects. The evaluations also demonstrated that a large quantity of training relations, or a high number of sentences from the corpus where individual relations match, positively impact the evaluated metrics.

The experiments revealed substantial differences in performance between individual predicates. Predicates that caused few disagreements between domain experts when manually labeling training relations perform better. The same observation holds for predicates which include clearly specified and tight domain, range and property restrictions. For some predicates (e.g. study), the presented algorithms reached first guess correct results of around 90% when choosing relation labels from ten alternatives.

A comparison of evaluation results between the methods presented in the thesis and two baseline scores illustrates highly significant gains in performance. For this purpose a random baseline and a baseline adopted from the literature were used.

The experiments also demonstrated the significant benefits achieved by the integration of knowledge inferred from external structured sources in the relation labeling process, as compared to relying on corpus-based methods only. However, current online datasets and ontologies involve certain data quality issues outlined in Section 5.3 (e.g. DBpedia redirects such as activist → activism, resulting in wrong concept grounding). But the advantages of incorporating external sources will increase over the next years, with as more and more linked data being made available online.

**Main Contributions.** In summary, the thesis contributes to compiling a common body of knowledge and advancing the state of the art by:

- Introducing a novel approach for the ontology learning task of labeling non-taxonomic relations. The thesis demonstrates the accuracy of the approach to learn specific relations, and compares it to state-of-the-art
techniques (although the various methods are not directly comparable due to different evaluation methodologies, numbers of relations to learn, underlying datasets, etc.).

- Integrates knowledge collected from Semantic Web sources with a machine learning approach, and a hybrid method that yields superior performance compared to corpus-based techniques alone.

- Providing a novel technique for linking concepts from a domain ontology to a meta ontology to gain additional semantics for supporting the ontology learning process.

- Implementing a modular and extensible architecture for generating relation label suggestions, independent of the application domain. Key modules such as the component for concept grounding can be adopted for other ontology learning tasks.

- Formally evaluating the overall architecture and its major components to assess the performance of the method. The experiments include a large training base of manually constructed training relations, a comprehensive domain corpus, and re-usable fragments of a meta ontology. The evaluation of Scarlet [47] demonstrates the benefits of hybrid approaches and current problems with methods solely relying on available Semantic Web data.

- Presenting a thorough overview of related work with a focus on relation labeling in ontologies.

**Future Research.** Despite these advances, there are a number of open issues that will require further attention. The following paragraphs outline major lines of *future work* to tackle some of these issues.

On the one hand, the current implementation assumes exactly one relation label to be appropriate for the relation between two concepts. Future research will investigate the implications of allowing multiple labels per relation. On the other hand, there are cases where none of the predefined labels is suitable, thresholds on similarity values will detect such situations. The author plans to determine the performance impacts and other consequences of raising or reducing the number of predefined predicates, as well as to apply the relation labeling architecture in other domains.

Several ideas have come up in the course of the present work on how to improve the performance of the corpus-based methods:
• Instead of only using verbs (and eventually prepositions) in verb vectors, the vector space model could include additional features such as co-occurring nouns, or even constituents such as (aggregated forms of) part-of-speech tags. Improved ways to capture context, e.g. applying a shallow parser, might enhance the distinction of relevant tokens from noise.

• A very interesting finding made during the evaluations of individual predicates was that passive-voice forms of predicates often yielded inferior performance as compared to their active-voice counterparts. This can presumably be attributed to the lemmatization process executed on all verbs before vector building; future research will confirm or discard this assumption by omitting or replacing lemmatization.

• Equation 4.30 relies upon static weights to utilize the conformance of a suggested relation label to ontological restrictions. These weights were set in an ad-hoc fashion before starting the evaluations. Future work will provide mechanisms to automatically optimize the weights based on existing training relations, and update the weights according to the evolution of the knowledge base.

An important line of development focuses on the improvement of concept grounding. The integration of additional sources, e.g. the Wikipedia category system which is represented by the SKOS vocabulary and the YAGO [180] classification schema, will help to raise the methods' recall. If grounding still fails, methods for the acquisition of synonyms or term resolution procedures (such as [196]) should be integrated. Disambiguation techniques to find the appropriate meaning of a term in cases where DBpedia returns disambiguation pages for input concept labels will address a similar problem. With the availability of additional structured data, advanced conflict resolution and mediation techniques will become an essential component of the refined grounding strategies.

For the practical application of the proposed methods it is crucial to reduce human effort involved in creating training relations and ontological definitions. Future work will comprise bootstrapping techniques (e.g. [52]) to support the automatic creation of training relations for particular relation types (predicates). Instead of defining domain and range restrictions manually, either existing specifications should be re-used, or mechanisms applied to learn the restrictions from existing training relations. After grounding concepts from training relations, the system can detect the appropriate restrictions automatically. It is presumably more effective to use some probabilistic model to specify and to validate ontological restrictions if they are
learned automatically, because concept type information gained from concept grounding includes a certain amount of misclassification. Next to the construction of training relations and the definition of ontological restrictions in the classification meta ontology, the specification of links between concepts in external ontologies (for example OpenCyc) and concepts from the meta ontology still requires significant human effort. Ontology mapping techniques that exploit lexical similarity could be used to automatically propose such links.