Supporting Lexical Ontology Learning by Relational Exploration*

Position Paper

Sebastian Rudolph, Johanna Völker, and Pascal Hitzler

Institute AIFB, Universität Karlsruhe, Germany
{sru,jvo,phi}@aifb.uni-karlsruhe.de

Abstract. Designing and refining ontologies becomes a tedious task, once the boundary to real-world-size knowledge bases has been crossed. Hence semi-automatic methods supporting those tasks will determine the future success of ontologies in practice. In this paper we describe a way for ontology creation and refinement by combining techniques from natural language processing (NLP) and formal concept analysis (FCA). We point out how synergy between those two fields can be established thereby overcoming each other's shortcomings.

1 Introduction

Along with the evolving Semantic Web, the need for elaborated techniques for generating and employing both large and complex ontologies emerges.

Beyond the "toy-examples" mostly used in research for demonstrating and investigating the basic principles for knowledge representation, the size of knowledge bases needed in real world applications will easily exceed the capabilities of human ontology designers to completely model a domain in an undirected, manual, ad-hoc manner.

Apart from the development of suitable methodologies for ontology creation and maintenance (which we consider an important topic but will not focus on in this paper), another way of assisting the modelling process is to provide semi-automatic methods which both

– intelligently suggest the extraction of potential knowledge elements (domain axioms / facts) from certain resources such as domain relevant text corpora and

* This work is supported by the German Federal Ministry of Education and Research (BMBF) under the SmartWeb project (grant 01 IMD01 B), by the Deutsche Forschungsgemeinschaft (DFG) under the ReaSem project, and by the European Union under the SEKT project (IST-2003-506826) and the NeOn project (IST-2005-027595).

– provide guidance during the knowledge specification process by asking decisive questions in order to clarify still undefined parts of the knowledge base.

Obviously, those two requirements complement each other. The first one clearly falls into the area of NLP. By using existing methods for knowledge extraction from texts, passages can be identified which indicate the validity of certain pieces of knowledge. For the second requirement, strictly logic-based exploration techniques are needed which yield logically crisp propositions. We argue that integrating these two directions of knowledge acquisition in one scenario will help overcoming disadvantages of either approach and complement each other.

In order to make this concrete, we will introduce Text2Onto (a tool for lexical ontology learning) and Relational Exploration (a dialogue-based method for knowledge acquisition) and sketch a way to combine those two approaches in a synergetic way.

2 Text2Onto – NLP techniques for ontology learning

Text2Onto [1] is a framework for ontology learning, i.e. the automatic acquisition of ontologies from textual data by natural language processing and machine learning techniques. It relies upon an ontology metamodel [2] which allows for attaching uncertainty values and provenance information to all ontology modeling primitives. Text2Onto features algorithms for generating concepts, taxonomic and non-taxonomic relationships as well as disjointness axioms [3] based on lexical evidence.

Clearly, it is advantageous to gain suggestions for constructing or refining the ontology based on textual data: though not fully automatized (since the last decision whether a detected piece of knowledge is really valid has to be left to the ontology engineer), a lot of textual search and estimation work can be shifted to algorithms leaving merely decisions to the expert.

Yet, there is no guarantee, that the automatically generated knowledge model is correct, and precise enough for characterizing the domain in question. This is because on one hand, there might be valid and also relevant pieces of knowledge present in the text which are not properly extracted by Text2Onto. And on the other hand, the corpus itself might not contain all valid domain knowledge to the wanted extent of precision.

Exploration techniques are good means to overcome the lack of completeness and precision in learned ontologies. It is just natural to apply them in order to further specify the knowledge beyond the information extractable from the corpus, which makes relational exploration a perfect complement to automatic approaches for ontology generation based on lexical resources.

3 Relational Exploration

The technique of Relational Exploration (short: RE, introduced in [4] and thoroughly treated in [5]) is based on the well-known attribute exploration algorithm (see [6, 7]) from formal concept analysis [8]. This algorithm is extended to a setting with unary and binary relations and uses description logic (DL) concept descriptions of the description logic $\mathcal{FLE}$ (see [9] for a comprehensive treatise on description logics) instead of "logically opaque" attributes. Hence it is possible to explore DL axioms (more precisely: general concept inclusion axioms, short: GCIs) with this techniques. I.e., in an interview-like process, a domain expert has to judge, whether a proposed GCI is valid in the domain he is describing. By employing a DL reasoner it can be furthermore guaranteed that during the exploration process, no redundant questions will be asked. The confirmed GCIs will then be added to the knowledge base thereby refining it. Since OWL DL [10] – the standard language for representing ontologies – is based on description logics, the RE method easily carries over to any kind of ontologies specified in that language.

The advantage of RE is that the obtained results are logically crisp and naturally consistent. Moreover, the refined knowledge base is even complete in the sense that any GCI formed out of $\mathcal{FLE}$ concept descriptions (of a certain role depth) that is valid in the described domain will be derivable from the refined knowledge base.

Yet, one major shortcoming of RE is the following: since the set of semantically different and possibly valid GCIs is growing rapidly with increasing role depth and number of atomic concept and roles, the number of asked questions will soon exceed the ontology designers resources.

4 Synergy and Conclusion

We will now sketch how an ontology refinement task can be accomplished by an intertwined application of Text2Onto and RE.

Suppose there is a knowledge base that has to be refined with respect to some specific term. Firstly, Text2Onto can be used to extract hypothetical axioms (subclass correspondences, concept instantiations, non-taxonomic relationships, and disjointness axioms) related to this term out of a document describing the domain of interest. These potential axioms will – after having been checked for validity by the expert – be added to the knowledge base.

In the next step, RE will be applied to further specify the interdependencies of the investigated concept and the (as indicated by Text2Onto's relevance measure) important related concepts. To speed up this process, some of the confirmed axioms can be directly provided to the exploration process as a-priori knowledge: simple subclass-of relations will be trivially encoded as attribute implications, concept instantiation will be stored as object-attribute-incidences in the underlying formal context, and the disjointness of the classes, say, $A_1, \ldots, A_n$ will be encoded as implications of the form $A_1, \ldots, A_n \to \bot$. During the subsequent actual exploration process, Text2Onto will be further applied in the following way: If the RE algorithm comes up with a hypothetical axiom, Text2Onto will look for possibly relevant passages in the corpus, pinpoint to them indicating their relevance, and possibly suggest answers (and assigned possibilities) to the expert.

To put the synergy in a nutshell: Exploration helps elaborating underspecified parts of textually acquired ontologies whereas text mining can contribute to restrict the exploration space and suggest answers to single design questions brought up as the exploration goes on.

The framework sketched in this position paper still has to be evaluated for its feasibility. Therefore, we are currently implementing a prototype to integrate relational exploration with our ontology learning tools, including Text2Onto and LExO [11], which will allow us to carry out first evaluation experiments, and to obtain empirical data in the near future.

References

1. Cimiano, P., Völker, J.: A framework for ontology learning and data-driven change discovery. In: Proc. of the NLDB'2005. (2005)
2. Vrandecic, D., Völker, J., Haase, P., Tran, D.T., Cimiano, P.: A metamodel for annotations of ontology elements in owl dl. In Sure, Y., Brockmans, S., Jung, J., eds.: Proceedings of the 2nd Workshop on Ontologies and Meta-Modeling, Karlsruhe, Germany, GI Gesellschaft für Informatik (2006)
3. Haase, P., Völker, J.: Ontology learning and reasoning - dealing with uncertainty and inconsistency. In da Costa, P.C.G., Laskey, K.B., Laskey, K.J., Pool, M., eds.: Proceedings of the Workshop on Uncertainty Reasoning for the Semantic Web (URSW). (2005) 45–55
4. Rudolph, S.: Exploring relational structures via FLE. In Wolff, K.E., Pfeiffer, H.D., Delugach, H.S., eds.: Conceptual Structures at Work: 12th International Conference on Conceptual Structures. Volume 3127 of LNCS., Huntsville, AL, USA, Springer (2004) 196 – 212
5. Rudolph, S.: Relational Exploration - Combining Description Logics and Formal Concept Analysis for Knowledge Specification. Universitätsverlag Karlsruhe (2006) Dissertation.
6. Ganter, B.: Algorithmen zur formalen Begriffsanalyse. In Ganter, B., Wille, R., Wolf, K.E., eds.: Beiträge zur Begriffsanalyse, B.I. Wissenschaftsverlag, Mannheim (1987) 241 – 254
7. Ganter, B.: Attribute exploration with background knowledge. Theoretical Computer Science **217** (1999) 215 – 233
8. Ganter, B., Wille, R.: Formal Concept Analysis: Mathematical Foundations. Springer-Verlag New York, Inc., Secaucus, NJ, USA (1997)
9. Baader, F., Calvanese, D., McGuinness, D.L., Nardi, D., Patel-Schneider, P.F., eds.: The Description Logic Handbook: Theory, Implementation, and Applications. In Baader, F., Calvanese, D., McGuinness, D.L., Nardi, D., Patel-Schneider, P.F., eds.: Description Logic Handbook, Cambridge University Press (2003)
10. McGuinness, D.L., van Harmelen, F.: OWL Web Ontology Language Overview. `http://www.w3.org/TR/2004/REC-owl-features-20040210/` (2004)
11. Völker, J., Hitzler, P., Cimiano, P.: Acquisition of OWL DL axioms from lexical resources. In: Proceedings of the 4th European Semantic Web Conference (ESWC'07), Springer (2007)