Foundations of Refinement Operators for Description Logics

Jens Lehmann\textsuperscript{1,*} and Pascal Hitzler\textsuperscript{2,**}

\textsuperscript{1} Universität Leipzig, Department of Computer Science
Johannisgasse 26, D-04103 Leipzig, Germany
lehmann@informatik.uni-leipzig.de

\textsuperscript{2} Universität Karlsruhe (TH), AIFB Institute
D-76128 Karlsruhe, Germany
hitzler@aifb.uni-karlsruhe.de

Abstract. In order to leverage techniques from Inductive Logic Programming for the learning in description logics (DLs), which are the foundation of ontology languages in the Semantic Web, it is important to acquire a thorough understanding of the theoretical potential and limitations of using refinement operators within the description logic paradigm. In this paper, we present a comprehensive study which analyses desirable properties such operators should have. In particular, we show that ideal refinement operators in general do not exist, which is indicative of the hardness inherent in learning in DLs. We also show which combinations of desirable properties are theoretically possible, thus providing an important step towards the definition of practically applicable operators.

1 Introduction

With the advent of the Semantic Web and Semantic Technologies, ontologies are becoming one of the most prominent paradigms for knowledge representation and reasoning. However, recent progress in the field faces a lack of available ontologies due to the fact that engineering such ontologies constitutes a considerable investment of resources. Methods for the automated acquisition of ontologies are therefore required. In this article, we develop theoretical foundations for the creation of such methods.

In 2004, the World Wide Web Consortium (W3C) recommended the Web Ontology Language OWL\textsuperscript{1} as a standard for modelling ontologies on the Web. In the meantime, many studies and applications using OWL have been reported in research, many of which go beyond Internet usage and employ the power of

\textsuperscript{*} The first author acknowledges support by the German Federal Ministry of Education and Research (BMBF) under the SoftWiki project (grant 01 ISF02 B).

\textsuperscript{**} The second author acknowledges support by the German Federal Ministry of Education and Research (BMBF) under the SmartWeb project (grant 01 IMD01 B) and by the Deutsche Forschungsgemeinschaft (DFG) under the ReaSem project.

\textsuperscript{1} http://www.w3.org/2004/OWL/
ontological modelling in other fields like software engineering, knowledge management, and cognitive systems.

In essence, OWL coincides with the description logic $\mathcal{SHOIN}(D)$ and is thus a knowledge representation formalism based on first-order logic. In order to leverage machine-learning approaches for the acquisition of OWL ontologies, it is therefore required to develop methods and tools for the learning in description logics. To date, only few investigations have been carried out on this topic, which can be attributed to the fact that description logics (DLs) have only recently become a major paradigm in knowledge representation and reasoning.

In this paper, we investigate the applicability of methods from Inductive Logic Programming (ILP) for the learning in description logic knowledge bases. We are motivated by the success of ILP methods and believe that similar results can be achieved for DLs.

Central to the usual ILP approach are the so-called refinement operators which are used to traverse the search space, and many approaches indeed hinge on the definition of a suitable such operator. Theoretical investigations on ILP refinement operators have identified desirable properties for them to have, which impact on their performance. These properties thus provide guidelines for the definition of suitable operators. It turns out, however, that for hard learning settings there are theoretical limitations on the properties a refinement operator can have. A corresponding general analysis therefore provides a clear understanding of the difficulties inherent in a learning setting, and also allows to derive directions for researching suitable operators.

In this paper we therefore give a full analysis of properties of refinement operators for description logics. To the best of our knowledge, such a complete analysis has not been done before, but the need for this investigation was expressed in [6,7]. The main contribution of this article is to derive a fundamental theorem about properties of refinement operators in DLs. This can serve as the foundation for the design of concrete refinement operators, which are used for induction – with potential applications in other areas of Machine Learning like clustering and data mining.

The paper is structured as follows. In Section 2 we will give a brief introduction to description logics. Section 3 formally describes the learning problem in description logics. Refinement operators and their properties are introduced. The main section is Section 4 which contains the results we obtained. We will fully analyse all combinations of interesting properties of refinement operators. This means we will show which combinations of properties are possible, i.e. for which combinations a refinement operator with these properties exists. We make only basic assumptions with respect to the description logic we are looking at, to cover as many description logics as possible. In Section 5 we discuss related work, in particular the relation to refinement operators in the area of traditional Inductive Logic Programming. Finally, in Section 6 we summarise our work and draw conclusions.

Some proofs are omitted for lack of space. They can be found in a separate technical report [12].