

Chapter 5

Learning of OWL Class Expressions on Very Large Knowledge Bases and its Applications

Sebastian Hellmann

Universität Leipzig, Germany

Jens Lehmann

Universität Leipzig, Germany

Sören Auer

Universität Leipzig, Germany

ABSTRACT

The vision of the Semantic Web aims to make use of semantic representations on the largest possible scale - the Web. Large knowledge bases such as DBpedia, OpenCyc, and GovTrack are emerging and freely available as Linked Data and SPARQL endpoints. Exploring and analysing such knowledge bases is a significant hurdle for Semantic Web research and practice. As one possible direction for tackling this problem, the authors present an approach for obtaining complex class expressions from objects in knowledge bases by using Machine Learning techniques. The chapter describes in detail how to leverage existing techniques to achieve scalability on large knowledge bases available as SPARQL endpoints or Linked Data. The algorithms are made available in the open source DL-Learner project and this chapter presents several real-life scenarios in which they can be used by Semantic Web applications.

INTRODUCTION

The vision of the Semantic Web aims to make use of semantic representations on the largest possible

scale - the Web. We currently experience that Semantic Web technologies are gaining momentum and large knowledge bases such as DBpedia (Auer et al., 2007), OpenCyc (Lenat, 1995), GovTrack (Tauberer, 2008) and others are freely available. These knowledge bases are based on semantic

DOI: 10.4018/978-1-60960-593-3.ch005

knowledge representation standards like RDF and OWL. They contain hundred thousands of properties as well as classes and an even larger number of facts and relationships. These knowledge bases and many more (ESWiki, 2008) are available as Linked Data (Berners-Lee, 2006; Bizer, Cyganiak, & Heath, 2007) or SPARQL endpoints (Clark, Feigenbaum, & Torres, 2008).

Due to their sheer size, users of these knowledge bases, however, are facing the problem, that they can hardly know which identifiers are used and are available for the construction of queries. Furthermore, domain experts might not be able to express their queries in a structured form at all, but they often have a very precise imagination what kind of results they would like to retrieve. A historian, for example, searching in DBpedia for ancient Greek law philosophers influenced by Plato can easily name some examples and if presented a selection of prospective results he will be able to quickly identify false results. However, he might not be able to efficiently construct a formal query adhering to the large DBpedia knowledge base a priori.

The construction of queries asking for objects of a certain kind contained in an ontology, such as in the previous example, can be understood as a class construction problem: We are searching for a class expression which subsumes exactly those objects adhering to our informal query (e.g. ancient Greek law philosophers influenced by Plato). Recently, several methods have been proposed for constructing ontology classes by means of Machine Learning techniques from positive and negative examples (Lehmann & Hitzler, 2007a, 2007b, 2010). These techniques are tailored for small and medium size knowledge bases, while they cannot be directly applied to large knowledge bases (such as the initially mentioned ones) due to their dependency on reasoning methods. In this paper, we present an approach for leveraging Machine Learning algorithms for learning of ontology class expressions in large knowledge bases, in particular those available as SPARQL (Clark et al.,

2008) endpoints or Linked Data. The scalability of the algorithms is ensured by reasoning only over “interesting parts” of a knowledge base for a given task. As a result users of large knowledge bases are empowered to construct queries by iteratively providing positive and negative examples to be contained in the prospective result set.

Overall, we make the following contributions:

- development of a flexible method for extracting relevant parts of very large and possibly interlinked knowledge bases for a given learning task,
- thorough implementation, integration, and evaluation of these methods in the DL-Learner framework (Lehmann, 2009)
- presentation of several application scenarios and examples employing some of the largest knowledge bases available on the Web.

This article is a revised and extended version. Since the original submission, several applications were created based on the presented method. These include the following tools (presented in the Section “Applications”):

- ORE (Lehmann & Bühmann 2010)
- HANNE (Hellmann et al. 2010)
- The Tiger Corpus Navigator (Hellmann et al. 2010)

In this article, we will first cover preliminaries, namely a quick introduction in Description Logics, OWL, and the learning problem we consider. We then briefly describe the underlying learning algorithm and present in detail how it can be applied on very large knowledge sources. In the next sections, we will mention some possible usage scenarios including some applications, which have been realized. We describe and evaluate our approach in several scenarios in the subsequent sections and, finally, conclude with some related work and an outlook on future work.

25 more pages are available in the full version of this document, which may be purchased using the "Purchase" button on the product's webpage:
www.igi-global.com/chapter/learning-owl-class-expressions-very/55043

Related Content

Embracing the Social Web for Managing Patterns

Pankaj Kamthan (2010). *Handbook of Research on Web 2.0, 3.0, and X.0: Technologies, Business, and Social Applications* (pp. 733-747).

www.igi-global.com/chapter/embracing-social-web-managing-patterns/39202

In Defense of Ambiguity

Patrick J. Hayes and Harry Halpin (2008). *International Journal on Semantic Web and Information Systems* (pp. 1-18).

www.igi-global.com/article/defense-ambiguity/2847

Can Knowledge Management Assist Firms to Move from Traditional to E-Commerce: The Case of Greek Firms

Irene Samanta (2009). *Social Web Evolution: Integrating Semantic Applications and Web 2.0 Technologies* (pp. 102-115).

www.igi-global.com/chapter/can-knowledge-management-assist-firms/29291

Advancing Learning Through Virtual Worlds

Steve Mahaley and Robin Teigland (2010). *Handbook of Research on Web 2.0, 3.0, and X.0: Technologies, Business, and Social Applications* (pp. 556-572).

www.igi-global.com/chapter/advancing-learning-through-virtual-worlds/39191