Chapter 7
Reasoning in Semantic Web-based Systems

Florian Fischer and Gulay Unel

Abstract In this chapter, we introduce the basics of formal languages and reasoning in a Web context. Denoting information by means of a logical formalism makes it possible to employ established techniques from the field of automated reasoning. However, reasoning in the context of Web-based systems has a distinct set of requirements in terms of quality and quantity of the information it has to cope with. In turn, this chapter focuses, in a very foundational way, on reasoning on Semantic Web-oriented data. For this purpose, we briefly identify and describe the basic paradigms forming the background for knowledge representation in Web-based Systems. We then examine how these paradigms are reflected in current standards and trends on the Web and what kinds of reasoning they typically facilitate. Based on this, we proceed to focus on concrete reasoning techniques and their particular properties, including optimizations and various other possibilities, e.g., parallelization and approximation, to meet the scalability requirements in Web-based systems.

7.1 Introduction

Web data grows faster than the capacity of Web-based systems that perform computation on this data. This growth also results in additional requirements for systems to go beyond keyword searches, use of document connectivity and usage patterns. Consider the following question:

Example 7.1 Find the names of the institutions who are involved in Semantic Web research via joint papers in related conferences.

An interpretation of additional data, i.e., metadata, that describes the contents of accessible resources is required to answer questions such as that given in Example 7.1. For this example, the relevant aspects of the natural language content of a document must be described through metadata.

The Semantic Web standardization activity of the World Wide Web Consortium (W3C) aims to design the basic technical standards that allow the representation of...
metadata on the Web, which in turn enables applications to derive answers to questions similar to Example 7.1. These standards can be seen as cornerstones for the development of future large-scale Semantic Web applications that primarily focus on two main aspects: machine readability and machine interpretability [45].

Machine readability requires the definition of a framework for the description of resources on the Web, i.e., the Resource Description Framework (RDF) [22]. RDF can be understood as a semi-structured data model for metadata on the Web. A knowledge-base of Web data typically contains metadata from a collection of information sources, e.g., several Web pages where these information sources are described using the RDF data model.

Machine interpretability is another key aspect for answering questions similar to Example 7.1 on a Web knowledge-base. For instance, information consumers and providers need to agree on the denotation of a common vocabulary, i.e., for Example 7.1, proper nouns such as “Semantic Web”, concepts such as “institution” and “conference” as well as the relations such as “involved”, must be represented in a structured way to make the evaluation of such queries possible. Web ontology languages allow for the description of this vocabulary by means of logical axioms. Ontology languages, including the RDF vocabulary description language (RDFS) [17], Web Ontology Language (OWL) [26], and the Web Service Modeling Language (WSML) [46], have been standardized for this purpose and are grounded on the assignment of formal semantics to their language constructs. This allows machines to interpret information such that new implicit facts can be derived from asserted facts using a form of symbolic processing, i.e., logical entailment.

In this chapter, we introduce the basics of formal languages and reasoning in a Web context. In Sect. 7.2, we formally introduce the relevant background in particular fundamentals of Logic Programming and Description Logics. Section 7.3 introduces some of the relevant standards for the Semantic Web, namely OWL 2 and WSML. Section 7.4 introduces the most common reasoning techniques that are used for the underlying paradigms introduced in Sect. 7.2. Section 7.5 examines the strategies for scalable reasoning in the context of the Web. Finally, conclusions and future research directions are given in Sect. 7.6.

7.2 Background

This section focuses on the underlying paradigms at a high-level. The content here is not specific to a concrete language.

7.2.1 Logic Programming

This subsection introduces Logic Programming [23] as a knowledge representation paradigm and especially focuses on the different ways to define its semantics.