

Chapter 7.19

Scalable Authoritative OWL Reasoning for the Web

Aidan Hogan

National University of Ireland, Ireland

Andreas Harth

National University of Ireland, Ireland

Axel Polleres

National University of Ireland, Ireland

ABSTRACT

In this article the authors discuss the challenges of performing reasoning on large scale RDF datasets from the Web. Using ter-Horst's pD* fragment of OWL as a base, the authors compose a rule-based framework for application to web data: they argue their decisions using observations of undesirable examples taken directly from the Web. The authors further temper their OWL fragment through consideration of "authoritative sources" which counter-acts an observed behavior which we term "ontology hijacking": new ontologies published on the Web re-defining the semantics of existing entities resident in other ontologies. They then present their system for performing rule-based forward-chaining reasoning which they call SAOR: Scalable Authoritative OWL Reasoner. Based upon observed characteristics

of web data and reasoning in general, they design their system to scale: the system is based upon a separation of terminological data from assertional data and comprises of a lightweight in-memory index, on-disk sorts and file-scans. The authors evaluate their methods on a dataset in the order of a hundred million statements collected from real-world Web sources and present scale-up experiments on a dataset in the order of a billion statements collected from the Web.

INTRODUCTION

Information attainable through the Web is unique in terms of scale and diversity. The Semantic Web movement aims to bring order to this information by providing a stack of technologies, the core of which is the Resource Description Framework

(RDF) for publishing data in a machine-readable format: there now exists millions of RDF data-sources on the Web contributing billions of statements. The Semantic Web technology stack includes means to supplement instance data being published in RDF with ontologies described in RDF Schema (RDFS) (Brickley and Guha 2004) and the Web Ontology Language (OWL) (Bechhofer, van Harmelen et al. 2004; Smith, Welty et al. 2004), allowing people to formally specify a domain of discourse, and providing machines a more sapient understanding of the data. In particular, the enhancement of assertional data (i.e., instance data) with terminological data (i.e., structural data) published in ontologies allows for deductive reasoning: i.e., inferring implicit knowledge.

In particular, our work on reasoning is motivated by the requirements of the Semantic Web Search Engine (SWSE) project: <http://swse.deri.org/>, within which we strive to offer search, querying and browsing over data taken from the Semantic Web. Reasoning over aggregated web data is useful, for example: to infer new assertions using terminological knowledge from ontologies and therefore provide a more complete dataset; to unite fractured knowledge (as is common on the Web in the absence of restrictive formal agreement on identifiers) about individuals collected from disparate sources; and to execute mappings between domain descriptions and thereby provide translations from one conceptual model to another. The ultimate goal here is to provide a “global knowledge-base”, indexed by machines, providing querying over both the explicit knowledge published on the Web and the implicit knowledge inferable by machine. However, as we will show, complete inferencing on the Web is an infeasible goal, due firstly to the complexity of such a task and secondly to noisy web data; we aim instead to strike a compromise between the above goals for reasoning and what is indeed feasible for the Web.

Current systems have had limited success in exploiting ontology descriptions for reasoning over RDF web data. While there exists a large body of work in the area of reasoning algorithms and systems that work and scale well in confined environments, the distributed and loosely coordinated creation of a world-wide knowledge-base creates new challenges for reasoning:

- the system has to perform on web-scale, with implications on the completeness of the reasoning procedure, algorithms and optimisations;
- the method has to perform on collaboratively created knowledge-bases, which has implications on trust and the privileges of data publishers.

With respect to the first requirement, many systems claim to inherit their scalability from the underlying storage—usually some relational database system—with many articles having been dedicated to optimisations on database schemata and access; c.f. (Pan and Heflin 2003; Theoharis, Christophides et al. 2005; Zhou, Ma et al. 2006; Hondjack, Pierra et al. 2007). With regards the second requirement, there have been numerous papers dedicated to the inter-operability of a small number of usually trustworthy ontologies; c.f. (Ghilardi, Lutz et al. 2006; Lutz, Walther et al. 2007; Jiménez-Ruiz, Grau et al. 2008). We leave further discussion of related work to Section 6, except to state that the combination of web-scale and web-tolerant reasoning has received little attention in the literature and that our approach is novel.

Our system, which we call “Scalable Authoritative OWL Reasoner” (SAOR), is designed to accept as input a web knowledge-base in the form of a body of statements as produced by a web-crawl and to output a knowledge-base enhanced by forward-chaining reasoning over a given fragment of OWL. In particular, we choose forward-chaining to avoid the runtime complexity

42 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/scalable-authoritative-owl-reasoning-web/37736

Related Content

Application of Big Data in Entrepreneurship and Innovation Education for Higher Vocational Teaching

Long Chen and Jiang He (2023). *International Journal of Information Technology and Web Engineering* (pp. 1-16).

www.irma-international.org/article/application-of-big-data-in-entrepreneurship-and-innovation-education-for-higher-vocational-teaching/333898

Cloud-Based Healthcare Systems: Emerging Technologies and Open Research Issues

Ahmed Shawish and Maria Salama (2016). *Web-Based Services: Concepts, Methodologies, Tools, and Applications* (pp. 2299-2323).

www.irma-international.org/chapter/cloud-based-healthcare-systems/140901

Web-Based Geospatial Services: Implementing Interoperability Specifications

Iftikhar U. Sikder, Aryya Gangopadhyay and Nikhil V. Shampur (2010). *Web Engineering Advancements and Trends: Building New Dimensions of Information Technology* (pp. 281-297).

www.irma-international.org/chapter/web-based-geospatial-services/40434

A Secure Data Transfer Approach With an Efficient Key Management Over Cloud

Lalit Mohan Gupta, Hitendra Garg and Abdus Samad (2022). *International Journal of Information Technology and Web Engineering* (pp. 1-21).

www.irma-international.org/article/a-secure-data-transfer-approach-with-an-efficient-key-management-over-cloud/306917

Neural Network for Big Data Sets

Vo Ngoc Phu and Vo Thi Ngoc Tran (2019). *Computational Intelligence in the Internet of Things* (pp. 271-303).

www.irma-international.org/chapter/neural-network-for-big-data-sets/224453