Parallelizing Semantic Web Inference using the Actor Model

Evan W. Patton

October 30, 2010

Abstract

Description logics are decidable logics that enable knowledge engineers to represent classifications of objects tractably. The Web Ontology Language (OWL) [1] has a description logic fragment to give Semantic Web [2] applications access to powerful logic while retaining elegant computational properties, but most work on OWL reasoners has focused on improving performance using a single thread. We expand on existing work in parallel reasoning [3] by using the Actor model [4] to represent resources on the semantic web and message passing between actors to initiate reasoning and communicate results.

1 Introduction

Scalability of reasoning on the semantic web is a major research area. OWL Full is undecidable and OWL-DL while being decidable, is still NP-complete. With data now ranging in the billions of triples, it is impossible to store the data on a single machine and reason over it. Therefore, more distributed systems will be necessary in order to reach the full potential of this new web and the data it provides. Up until recently, most work was focused on optimization of OWL-DL reasoners using tableau mechanisms [5].

Parallelization of OWL reasoning, being critical to the scalability of the semantic web, is an open research topic. Soma & Prasanna [3] have demonstrated that data partitioning could increase reasoning performance sublinearly, ranging anywhere from four to eighteen times the speed on a 16 core machine. They also demonstrated that partitioning the rule set for OWL, rather than the data, also resulted in sublinear increase in reasoning capabilities, although at a markedly lower rate than in the case of data partitioning. While this parallelization method was effective overall in improving reasoning performance, it still works entirely within a single, closed system.

2 Proposed Work

To expand the previous work, this work, instead of making specific cuts designed to solely optimize the processing on a multicore machine, will instead assume that individuals creating data on the semantic web do so in a manner that logically groups data related to a specific concept, and that these data are distributed across the web. The concept, having a URI, therefore, could be a logical partition of the data without requiring extensive machine analysis. It also allows for data to be stored anywhere on the web and agents simply post messages to a mailbox named by the resource's URI.

Consider the Wine Agent application [**?**], which provides an ontology for describing wines and food on the web. If a winery, XYZ Vineyards, posts information about its wines online as a static file (as is often the case), that information must be downloaded by different agents separately and they must reason about the information independently of one another, ultimately wasting many cycles computing the same set of information. If, however, the URIs identified by XYZ Vineyards were actors listening for requests from clients, it would be possible for one agent to submit a classification query about a wine to the actor that encapsulates all knowledge about that wine. This actor could then reply to the query and also the response so future classification queries from clients could be answered immediately instead of each client computing the result.

This work will not find superlinear increases in raw reasoning power due to the relatively high latency of the Internet compared to a cluster of computers. Instead, the distributed nature of semantic web applications will be able to take advantage of shared reasoning resources to reduce the amount of energy computation requires when considering the whole system. By the end of the semester, a demonstrable system which exposes the individuals and classes described in the Wine Agent's ontology and classifies them using multiple machines on the Internet should be completed.

References

[1] Motik, B., Patel-Schneider, P.F., & Parsia, B. (Eds.). (2009). "OWL 2 Web Ontology Language: Structural Specification and Functional-Style Syntax". `http://www.w3.org/TR/2009/REC-owl2-syntax-20091027/`

[2] Berners-Lee, T., Hendler, J., & Lassila, O. (2001). "The Semantic Web". *Scientific American Magazine*".

[3] Soma, R. & Prasanna, V.K. (2008). "Parallel Inferencing for OWL Knowledge Bases". In: *37th International Conference on Parallel Computing*.

[4] Agha, G.A., Mason, I.A., Smith, S.F., & Talcott, C.L. (1993). "A Foundation for Actor Computation". *Journal of Functional Programming 1*(**1**). Cambridge University Press.

[5] van Eijck, J., Heguiabehere, J., & Nuallin, B.. (2001). "Tableau reasoning and programming with dynamic first order logic". *Logic Journal of the IGPL, 9*(**3**).

[6] Patton, E.W. & McGuinness, D.L. (2009). "The Mobile Wine Agent: Pairing Wine with the Social Semantic Web". In: *2nd Social Data on the Web Workshop.*