Scalable Distributed OWL Reasoning for the Semantic Web

Introduction

The Semantic Web [1] is an extension of the traditional Web whereby data on the Web can have a meaning and hence can be processed by automated intelligent agents to assist humans. In Semantic Web the meaning of data is encoded in language such as RDF and OWL which are readable to both humans and machines.

In the Semantic Web machines can retrieve certain information more efficiently and be able to drive new information from existing data [1]. The process of driving new information from existing data is called reasoning and considering the amount of data available on the Web today the reasoning process is both data and computationally intensive. For example the Web Ontology Language (OWL) has three sublanguages OWL Full, OWL DL and OWL Lite. OWL full is un-decidable whilst OWL DL has a worst-case computational complexity of NExpTime [3]:

$$\mathrm{NEXPTIME} = \bigcup_{k \in \mathbb{N}} \mathrm{NTIME}(2^{n^k})$$

OWL Lite has a worst-case computational complexity of ExpTime [3]:

$$\mathrm{EXPTIME} = \bigcup_{k \in \mathbb{N}} \mathrm{DTIME}\left(2^{n^k}\right).$$

To overcome the computational complexity distributed systems can be used to provide some degree of parallelism and distribute the workload over a large number of computational resources. Data partitioning is one of the main challenges that faces distributed technique especially if the input data is highly correlated.

A number of distributed reasoning techniques have been established both in a Peer-to-Peer [5] and grid formation that can reason over a small data set [6]. However, most of these techniques perform poorly when presented with real world data, for example recent attempt by Urbani et al [6] to utilize the MapReduce distributed programming model [2] to reason over large Ontology data sets has shown poor OWL reasoning performance.

The aim of this research is to develop a distributed reasoning technique for OWL 2 [4] that is efficient, scalable and addresses the shortcomings of the currently available distributed OWL reasoners.

Research Approach and Methods

The objectives of this research are outlined as follows:

1- Explore and access the currently available distributed ontology reasoning techniques.
2- Identify the performance and scalability shortcomings with the available techniques and how these can be addressed by utilizing the current advances in Grid and Cloud computing.
3- Taking the above as input, develop a distributed reasoning technique for OWL 2.
4- Build a reasoning engine for the developed technique and assess its efficiency and scalability

Objectives 1 and 2 will be achieved by in-depth literature review and data analysis of published work. Objectives 3 and 4 will be achieved through theory development and practical laboratory work to develop a distributed reasoning framework and critically assess its capabilities.

Programme of Work

Staring in May 2010, the work will be conducted and organized in a series of work-packages as follows:

1- WP1: Literature review and theoretical development (9 months)
2- WP2: Identification of current state-of-art shortcomings and devising techniques to address them (9 months)
3- WP3: Implementation and build of the devised techniques into a practical reasoning framework (14 months).
4- WP4: Data analysis and critical assessment of the produced results (8 months)
5- WP5: Writing up and submission (8 months)

Outputs and Benefits

The research seeks to develop a scalable, efficient distributed reasoning technique and implement it as a reasoning framework. This will contribute to the fields of distributed automated reasoning for the Semantic Web and makes clearer the various approaches that can be taken to address the common reasoning problem in a distributed manner.

References

1- Berners-Lee, T., Hendler, J. and Lassila, O. (2001). *The Semantic Web*. Scientific American, 284(5): 34-43.
2- Dean, J. and Ghemawat, S. (2004). *MapReduce: Simplified data processing on large clusters*. In Operating Systems Design and Implementation, pages 137-150.
3- Hitzler, P., Krötzsch, M., and Rudolph, S. (2009). *Foundations of Semantic Web Technologies*. Boca Raton: CRC, 2010. Print.
4- Motik, B., Patel-Schneider P.F., and Parsia, B. (2009). OWL 2 Web Ontology Language: Structural Specification and Functional-Style Syntax. W3C Candidate Recommendation, 11 June 2009. Available at http://www.w3.org/TR/2009/CR-owl2-syntax-20090611/.
5- Serafini L. and Tamilin A. (2005). *DRAGO: Distributed reasoning architecture for the semantic web*. In Proc. of the Second European Semantic Web Conference (ESWC'05), Heraklion, Greece.
6- Urbani J., Kotoulas S., Oren E., and Harmelen F. (2009). Scalable distributed reasoning using MapReduce. In Proceedings of the International Semantic Web Conference (ISWC). Nov. 2009.