

Towards a Global Scale Semantic Web

Zhengxiang Pan

Department of Computer Science and Engineering, Lehigh University
19 Memorial Dr. West, Bethlehem, PA 18015, U.S.A.
zhp2@lehigh.edu

1 Introduction

By transforming the Web from a collection of documents to a collection of semantically rich data sources, the Semantic Web promises an unprecedented benefit of a global knowledge infrastructure. My PhD research will try to help make that happen at a global scale by doing the following proposed work:

1. Design and implement a highly scalable Semantic Web knowledge base system by exploiting modern relational database technologies combined with the state-of-the-art description logics reasoners.
2. Build and empirically verify a framework that handles ontology evolution and the reuse of data on top of the perspective theory [1]. This framework should be able to relieve inconsistency and heterogeneity in a global scale Semantic Web.
3. Systematically evaluate the resulting system's capability and scalability in processing, integrating and querying data under the real world environment.

2 Problem Statement

Traditionally Semantic Web knowledge bases are based on description logic reasoners. These reasoners usually employ tableaux algorithms to do reasoning. Since most of the inferences are performed in main memory, their scalability are limited by the size of available physical memory. There has been a growing interest in the development of systems that will store and process large amount of Semantic Web data. However, most of these systems are geared towards RDF and RDF(S) data and therefore focus less on OWL reasoning. The above two types of systems represent two conflicting requirements on knowledge bases: completeness and scalability. The trade-off between these must be decided by each application. Nevertheless, an approach that could scale well and keep essential reasoning powers is highly desired in the Semantic Web at global scale.

The Semantic Web at global scale is also dynamic, inconsistent and heterogeneous. A framework that can handle ontology evolution, including ontology extension and the reuse of data, should be able to relieve these issues by integrating various ontologies and data sources.

3 Methodology

My proposed research will combine relational database techniques and description logic reasoning services to archive high scalability with essential inference capabilities. Generally, TBox reasonings will be handled by the DL reasoners and some ABox reasonings

will be implemented using database techniques. The relational database management systems (RDBMS) will be used to persistently store processed Semantic Web data and its high availability and scalability will also be employed to support queries at large volumes.

Typical ABox reasoning includes “realization”, which is to infer the implicit class memberships for individuals. I believe some of the cases can be handled in the form of database queries. The feasibility of this method has been initially investigated in [2], except that they focused on loading the data in the opposite direction.

I will also construct a framework to support ontology evolution and data reuse. The framework will be built on top of ontology perspectives [1]. This framework could reduce the occurrence of inconsistency since only a relevant subset of the Semantic Web is involved in processing a query. It can also facilitate the integration of information resources by creating new perspectives that include different ontologies and data sources.

Unbiased Semantic Web benchmarks such as LUBM [3] will be employed to make and validate design decisions when building the system. Real Semantic Web data will be used to evaluate the performance of the system comparing to similar existing systems, including those use deductive databases.

My prior work includes DLDB, a knowledge base system that extends a RDBMS with additional capabilities for partial OWL reasoning. The queries supported by current DLDB fall under extensional conjunctive queries. Our experiments with DLDB have demonstrated encouraging scalability [3]. In a recent work of our lab [4], we have used Swoogle’s 2005 index as our dataset (45 million triples). We also showed that by adding the mapping ontologies and using them as query perspectives, we retrieve more results with a slight increase in query time.

4 Conclusion

I have proposed a PhD research in building a Semantic Web knowledge base system at global scale. This work faces a couple of challenges. First, the sheer size of Semantic Web poses a critical requirement on scalability. Existing techniques found in artificial intelligence area do not seem to be capable of handling that large scale; whereas existing techniques in database area do not seem to be capable of reasoning on that rich semantics. Second, the inherently inconsistent and heterogeneous Semantic Web makes it very difficult to answer queries in a complete and sound fashion.

References

1. Heflin, J., Pan, Z.: A model theoretic semantics for ontology versioning. In: Proc. of the 3rd International Semantic Web Conference. (2004) 62–76
2. Borgida, A., Brachman, R.J.: Loading data into description reasoners. In: SIGMOD Conference. (1993) 217–226
3. Guo, Y., Pan, Z., Heflin, J.: LUBM: A benchmark for owl knowledge base systems. *Journal of Web Semantics* 3(2) (2005) 158–182
4. Pan, Z., Qasem, A., Heflin, J.: An investigation into the feasibility of the semantic web. In: Proc. of the Twenty First National Conference on Artificial Intelligence (AAAI-06). (2006) 1394–1399