A LarKC Approach to Development of Service-Oriented Semantic Reasoning Applications

Alexey Cheptsov[1]

[1] High-Performance Computing Center Stuttgart, Nobelstr. 19,
70569 Stuttgart, Germany
cheptsov@hlrs.de

Abstract. Reasoning is one of the essential application areas of the modern Semantic Web. At present, the semantic reasoning algorithms are facing significant challenges when dealing with the emergence of the Internet-scale knowledge bases, comprising extremely large amounts of data. The traditional reasoning approaches have only been approved for small, closed, trustworthy, consistent, coherent and static data domains. As such, they are not well-suited to be applied in data-intensive applications aiming on the Internet scale. We introduce the Large Knowledge Collider as a platform solution that leverages the service-oriented approach to implement a new reasoning technique, capable of dealing with exploding volumes of the rapidly growing data universe, in order to be able to take advantages of the large-scale and on-demand elastic infrastructures such as high performance computing or cloud technology.

Keywords: Semantic Web, Incomplete Reasoning, LarKC, service architecture

1 Introduction

The large- and internet-scale data applications are the primary challenger for the modern Semantic Web, and in particular for reasoning algorithms, used for processing exploding volumes of data exposed currently on the Web. Reasoning is the process of making implicit logical inferences from the explicit set of facts or statements, which constitute the core of any knowledge base. The key problem for most of the modern reasoning engines such as Jena or Pellet is that they can not efficiently be applied for the real-life data sets that consist of tens, sometimes of hundreds of billions of triples (a unit of the semantically annotated information), which can correspond to several petabytes of digital information on the disc. Whereas modern advances in the Supercomputing domain allow this limitation to be overcome, the reasoning algorithms and logic need to be adapted to the demands of the rapidly growing data universe, in order to be able to take advantages of the large-scale and on-demand computing infrastructures. On the other hand, the algorithmic principals of the reasoning engines need to be reconsidered as well in order to allow for the specific of very large volumes of data (e.g. inconsistent and noisy data). Service-oriented architectures (SOA) can greatly contribute to this goal, acting as the main enabler of the newly proposed reasoning techniques such as incomplete reasoning.

2 The Large Knowledge Collider

One of the most prominent efforts to facilitate the development of trend-new applications for large-scale reasoning has been the EU-funded project of the Large Knowledge Collider (LarKC) [1]. The mission of the project was to set up a distributed reasoning infrastructure for the Semantic Web community, which should enable the application of reasoning to scale far beyond the currently recognized limitations, by implementing the incomplete reasoning approach. The current and future Web applications that deal with "big data" are in focus of LarKC.

To realize this mission, LarKC has created an infrastructure that allows construction of plug-in-based reasoning applications, following the incomplete reasoning approach, facilitated by incorporating interdisciplinary techniques such as inductive, deductive, interleaved reasoning, in combination with the methods from other knowledge representation domains such as information retrieval, machine learning, cognitive and social psychology, etc. The core of the infrastructure is a platform – a software framework that facilitates design, testing, and exploitation of new reasoning techniques for development of large-scale applications. The platform does this by providing solutions for creating very lightweight, portable and unified services for data sharing, accessing, transformation, aggregation, and inferencing, as well as building Semantic Web applications on top of those services. The efficiency of the LarKC services is ensured by providing a transparent access to the underlying resource layer, served by the platform, involving elastic high performance computing, storage, and cloud resources, and in the other way around, providing performance analysis and monitoring information back to the user. The platform is built in a distributed, modular, and open source fashion. Moreover, the platform offers means for building and running applications across the plug-ins, provide them a persistent data layer for storing data, facilitate parallel execution of large-scale data operations by leveraging the distributed and high-performance resources.

Guided by the preliminary goal to facilitate the incomplete reasoning, LarKC has evolved in a unique platform, which can be used for the development of robust, flexible, and efficient semantic web applications, also leveraging the modern grid and cloud resources. Our poster presents the LarKC approach to developing the service-oriented reasoning applications. We demonstrate some of the most successful applications developed on the LarKC foundation such as Bottari – the Semantic Challenge winner in 2011, or WebPIE – the Billion Triple Challenge winner in 2010. We show that LarKC can be applied to solve many of the current Semantic Web challenges, and would like to serve a discussion forum on its adoption in the external development communities.

References

1. M. Assel, A. Cheptsov, G. Gallizo, I. Celino, D. Dell'Aglio, L. Bradeško, M. Witbrock, E. Della Valle: Large knowledge collider: a service-oriented platform for large-scale semantic reasoning. Proceedings of the International Conference on Web Intelligence, Mining and Semantics (WIMS'2011)