S

ScaleSem Approach to Check and to Query Semantic Graphs

Mahdi Gueffaz
University of Burgundy, France

Sylvain Rampacek
University of Burgundy, France

Christophe Nicolle
University of Burgundy, France

INTRODUCTION

Ontologies play an important role in the Semantic Web (Berners-Lee and al., 2001), and provide the basis for the definition of concepts and relationships that make information integration possible. Knowledge represented in ontologies can be used to annotate data, distinguish similar concepts, and generalize and specialize concepts published by data sources or produced by Web services. A great number of ontologies have become available; some of these ontologies can be very large, with impact on the tasks of ontology query answering and reasoning.

The number of available data sources and services has exploded in the web during the last few years. The ontology construction goes through several stages. Among them, the evolution step change consists in making the ontology more accurate and appropriate to the domain. Ontology evolution is a critical task because the new implementation can lead to the apparition of incoherences. Ontology inconsistency can occur for several reasons such as: modeling errors when correcting or adapting the ontology domain, conceptualization or specification. Incoherence corresponds to the existence of an unsatisfiable concept in the ontology intension.

A survey of popular semantic graphs query languages conducted by the World Wide Web Consortium W3C identified more than 20 languages that are either under development or have been implemented (Angles & Gutierrez, 2005). Some along the lines of traditional database query languages (e.g. SQL (Structured Query Language)), others based on logic and rule languages.

Since 2004, SPARQL has been rapidly adopted as the standard for querying the Semantic Web data. The W3C SPARQL (Prudhommeaux & Seaborne, 2005) is a query language for semantic graphs designed to meet such requirements and defines a query language with a SQL-like style, where a simple query is based on query patterns, and query processing consists of binding of variables to generate pattern solutions (graph pattern matching). SPARQL is undergoing improvements and has a lot of limitations.

We introduce a new method to query and to check the semantic graphs by using the Model Checking technique in order to reduce the incoherence in these graphs and make the data more relevant. Model Checking is an automatic verification technique, it has been applied to many cases in industry, for example (Baier & Katoen, 2008), in the Netherlands, it has revealed several serious flaws in the design of control system of a barrier protection against flooding which protects the main port of Rotterdam against floods. It is a powerful tool for system verification, because it can reveal errors that were not discovered by other formal methods, such as testing or simulation. It uses temporal logic to describe the properties checking the system model.

Model Checking can handle complex problems with large amounts of information, stored as a graph, in order to verify critical systems. In comparison, in the semantic web, the use of graphs is pervasive and serious problems of scalability appear (Homma et al., 2009). Thus, it is appropriate to use the algorithms developed for Model Checking in the field of the Semantic Web. We have implemented a tool box called "ScaleSem" which contains tools to manipulate this semantic graph.

DOI: 10.4018/978-1-4666-5888-2.ch718

BACKGROUND

The Semantic Web expression, attributed to Tim Berners-Lee (Berners-Lee, 2007) within the W3C, primarily refers to the vision of the future Web as vast resource exchange between humans and machines for an operation, qualitatively superior, large volumes of information and various services. The Semantic Web aims at organizing and structuring the huge quantity of information present on the Net. It consists of a semi-structured language based on XML. The W3C represents the semantic web in several layers. Each layer is built upon the layers below it. Thus, the whole set of layers uses the XML syntax.

Several languages for the Semantic Web have been developed in recent years. The most important of them are RDF (Becket & McBride, 2004) and OWL (McGuinness & van Harmelen, 2004). RDF is used to represent semantic graphs corresponding to specific knowledge modeling. OWL is a knowledge representation language built on the RDF data model. It provides the means for defining structured web ontologies. In practice, OWL is designed as an extension of RDF language; OWL is intended for description of classes and property types. Therefore, it is more expressive than RDF, which some blame a lack of expressiveness due to the only definition of relations between objects by assertions. OWL also provides better and easier integration, evolution, sharing and inference of ontologies.

The Resource Description Framework (RDF) is a language developed by the W3C to bring a semantic layer to the Web. It allows the connection with the Web resources using directed labeled edges. The structure of the RDF documents is a complex directed labeled graph. An RDF document is a set of triples <subject, predicate, object> as shown in Figure 1.

- **Resources:** All data objects described by a RDF statement are called resources. For example, resources are web sites or books. A resource is identified by an URI (Uniform Resource Identifier).

Figure 1. RDF triple

- **Properties:** A specific aspect, characteristic or relation of a resource is described by a property. For example, properties are the creation date of a web site or the author of a book.
- **Statements:** A statement combines a resource with its describing property and the value of the property. RDF statements are the structural building blocks of the language.

The resources are represented by circles in the graph; the properties are represented by directed arcs, and the values by a box (a rectangle). Values can be resources if they are described by additional properties. For example, when a value is a resource in another triplet, the value is represented by a circle.

It is a well-known fact in the community of the Semantic Web that ontologies play a key role in the delivery of the Semantic Web, by facilitating information sharing among communities of humans and software agents. Ontology is based on the RDF graphs, and it is also considered like a semantic graph. The word ontology was taken from philosophy, where it means a systematic explanation of being. In the field of artificial intelligence, ontology is described to "define the terms and the basic relationships of a domain vocabulary and the rules that specify how to combine terms and relations so as to extend the vocabulary." (Neches et al., 1991) To represent ontologies, the W3C provides a standard, known as OWL (Ontology Web Language). OWL is currently built on RDF, and it adds the ability to define classes in more complex connectors corresponding to the description logic equivalent (intersection, union, various restrictions, etc....), disjoint classes, the inverse or transitive properties or even the cardinality restrictions on properties.

To handle the RDF graphs, several designs and implementations of RDF query languages have been proposed. In 2004, the RDF Data Access Working Group, part of the W3C Semantic Web Activity, released a first public working draft of a query language for RDF, called SPARQL (Liu et al., 2010). Since then, SPARQL has been rapidly adopted as the standard for querying the Semantic Web data. In January 2008, SPARQL became a W3C Recommendation. SPARQL queries (Chebotko et al., 2006) are pattern matching queries on triples that constitute an RDF data graph.

Formal methods (Baier & Katoen, 2008) offer great potential for an early inclusion of verification