Chapter 4
Structuring Abstraction to Achieve Ontology Modularisation

Zubeida Khan
Council for Scientific and Industrial Research, Pretoria, South Africa

C. Maria Keet
Department of Computer Science, University of Cape Town, South Africa

ABSTRACT

Large and complex ontologies lead to usage difficulties, thereby hampering the ontology developers' tasks. Ontology modules have been proposed as a possible solution, which is supported by some algorithms and tools. However, the majority of types of modules, including those based on abstraction, still rely on manual methods for modularisation. Toward filling this gap in modularisation techniques, the authors systematised abstractions and selected five types of abstractions relevant for modularisation for which they created novel algorithms, implemented them, and wrapped them in a GUI, called NOMSA, to facilitate their use by ontology developers. The algorithms were evaluated quantitatively by assessing the quality of the generated modules. The quality of a module is measured by comparing it to the benchmark metrics from an existing framework for ontology modularisation. The results show that the module's quality ranges between average to good, whilst also eliminating manual intervention.

INTRODUCTION

An ontology is a logic-based model that is used to represent a subject domain in a machine-processable language for Semantic Web applications. There are various formal definitions presented in the literature for an ontology. A comprehensive definition for an ontology is as follows: (Guarino, 1998):

An ontology is a logical theory accounting for the intended meaning of a formal vocabulary, i.e. its ontological commitment to a particular conceptualization of the world. The intended models of a logical

DOI: 10.4018/978-1-7998-6697-8.ch004

language using such a vocabulary are constrained by its ontological commitment. An ontology indirectly reflects this commitment (and the underlying conceptualization) by approximating these intended models.

Ontologies are commonly used for some purpose such as enhanced searches across many databases or decision-making support for problems. It consists of entities, organised in a hierarchy, with relations among them and axioms to define such relations. The entities are often constrained and described using the other entities to provide better meaning.

Ontologies that describe large, well-defined domains are consequently large and complex in nature; e.g., the NCBI Taxonomy with its 1,015,206 classes (Federhen, 2012), and likewise SNOMED CT (Donnelly, 2006) and the FMA (Rosse & Mejino, 2003) are well-known to be challenging due to their size. This leads to difficulties for tools and humans alike: tools cannot process them due to computational limitations while humans face cognitive overload for understanding them. Over the last few years, there has been a growth in using modularity to assist with large ontologies (Amato et al., 2015; d'Aquin et al., 2006; Del Vescovo, 2011; Grau et al., 2008; Khan & Keet, 2016b). A module has been defined as follows (Khan & Keet, 2015):

A module M is a subset of a source ontology O, M Ì O, either by abstraction, removal or decomposition, or module M is an ontology existing in a set of modules such that, when combined, make up a larger ontology. Module M is created for some use-case U, and is of a particular type T. T is classified by a set of annotation features P, and is created by using a specific modularization technique MT, and has a set of evaluation metrics EM which is used to assess the quality of module M.

The general idea of modularisation refers to dividing and separating the components of a large system such that it can be recombined into a whole. The purpose of modularity is used to simplify and downsize an ontology for the task at hand, i.e., to modularise a large ontology into smaller ontologies, and it is required in ontology development and use when one needs to hide or delete knowledge that is not required for the use-case (Studer, 2010). Modularisation has been applied to various ontologies to improve usability and assist with complexity; e.g., the myExperiment ontology (Newman et al., 2009), the Semantic Sensor Net ontology (Janowicz & Compton, 2010), several BioPortal ontologies (Del Vescovo et al., 2011), and the FMA ontology (Mikroyannidi et al., 2009).

There are several modularisation methods and tools (Amato et al., 2015; Chen et al., 2019; d'Aquin et al., 2006; Hamdi et al., 2010; Kalyanpur et al., 2006; LeClair et al., 2019). For ontologies that are created in a modular way, however, many of them are created by manual methods (Khan & Keet, 2015). There is tool support for graph partitioning (Kalyanpur et al., 2006), query-based (Natalya Fridman Noy & Musen, 2000) and locality-based (Grau et al., 2008) techniques. We zoom in on the type of modules obtained from abstractions, taken from an existing categorisation of ontology modules (Khan & Keet, 2015). Abstraction is the principle of simplifying complex models by removing some unnecessary details based on some criteria, such as reducing a class hierarchy's depth or removing axioms to fit the ontology in a language of lower expressiveness. The purpose of abstraction, like modularity, is to have a simplified version of an ontology for a specific task or application.

Seeing that automatic techniques for generating these types of modules are lacking (discussed below), we solve this problem by investigating and structuring proposed types of abstractions, which led to the creation of a basic theory from the related works. Based on the theory, several abstractions were selected to fill in the gaps for the lacking tool support and we propose new algorithms for generating