Ontology Analysis on Complexity and Evolution Based on Conceptual Model*

Zhe Yang, Dalu Zhang, and Chuan Ye

Department of Computer Science and Technology, Tongji University,
Postal Code 20 18 04, Shanghai, P.R.C
{Hatasen, Shtjjsjyjx}@hotmail.com,
Daluz@ieee.org

Abstract. With the tremendous development in size, the complexity of ontology increases. Thus ontology evaluation becomes extremely important for developers to determine the fundamental characteristics of ontologies in order to improve the quality, estimate cost and reduce future maintenance. Our research examines the concepts and their hierarchy in ontology conceptual model, the common feature of most ontologies, which reflects the fundamental complexity. We suggest some well-defined metrics of complexity, which mainly examine the quantity, ratio and correlativity of concepts and relationships, to evaluate ontology from the viewpoint of complexity and evolution. In the study, we measured three ontologies in Gene Ontology to verify our metrics. The results indicate that these metrics works well, and the biological process ontology is the most complex one from the view of complexity, and the molecular function ontology is the unsteadiest one from the view of evolution.

1 Introduction

Large standardized ontologies are often developed by several researchers in parallel, such as GO [1]; a number of ontologies grow in the context of peer-to-peer applications [2]; other ontologies are constructed dynamically [3]. Although it becomes important to determine fundamental characteristics of ontologies [4], there are still very few commonly agreed methodologies and metrics to analyze and evaluate ontology complexity and evolution [5, 6]. Thus, metrics are expected to help developers to design ontologies, improve quality, estimate and reduce future maintenance costs in all life cycle.

The rest of this paper is structured as following. Section 2 reviews some related works about ontology, evolution, evaluation and other metrics. Section 3 introduces some formal notations of ontology conceptual model and proposes some complexity metrics. In section 4, the complexity analysis results of Gene Ontology are given to show its evolution trend. Section 6 is summary and outlook for future works.

* Supported by National Natural Science Foundation of China under Grant No. 90204010.

U. Leser, F. Naumann, and B. Eckman (Eds.): DILS 2006, LNBI 4075, pp. 216 – 223, 2006.

2 Related Works

The variety of causes and consequences of the ontology changes makes ontology evolution a very complex operation that should be considered as both an organizational and a technical process [7].

In contrast to the researches on ontology evolution and versioning, only little empirical work has focused on evaluation [8]. Ontology metrics is desired in ontology evaluation. Although some metrics have been suggested [9], more work is needed [8]. The most existing metrics are proposed to evaluate the syntactic, semantic, and structure of ontology conceptual model. There are few metrics investigating the ontology complexity and evolution.

Burton et al. assessed the effectiveness of the DAML ontologies [10]. They suggested an ontology auditor metrics suite, and mainly considered the syntactic, semantic, pragmatic and social quality of ontologies.

Literature [11] proposed a set of ontology cohesion metrics to measure the modular relatedness of OWL ontologies. These metrics is focused on the number of classes and depth of inheritance tree of all classes. And it computes cohesion metrics conceptually based on predefined OWL primitives, which explicitly defined tree-based semantic hierarchies in OWL ontologies.

In literature [12], authors use weighted class dependence graphs to represent a class diagrams, and present a structure complexity measure for the UML class diagrams based on entropy distance. It considers complexity of both classes and relationships between the classes. This method can measure the structure complexity of class diagrams objectively.

Idris studied two conceptual integrity metrics based on graph theory is his PhD thesis [13], which are conceptual coherence and conceptual complexity. Conceptual coherence uses average distance between nodes in a graph to measure the interrelatedness of concepts. And conceptual complexity reflects the average number of relationships per node with the average degree across all nodes in a graph.

Chris Mungall researched the increased complexity of Gene Ontology [14]. He measured the average number of paths-to-top of a term and used the path-to-term ratio to measure of complexity in an ontology, which is represented in DAG (directed acyclic graph). However, in his calculation of the total number of terms, the obsolete terms does not be eliminated. While calculating the paths-to-top of terms, the paths of these obsolete terms are not counted. Thus his result is not correct.

3 Ontology Conceptual Model and Complexity Metrics

3.1 Common Formal Notation

We use the following notation to represent some terms in the ontology conceptual model. And small letters are used to identify the notations related to concepts and relationships, while capital letters are used to identify the terminology related to ontology and metrics.

$C = \{c_1, c_2, \cdots, c_m\}$: the set of m concepts defined in an ontology explicitly. In other ontologies, concept may be named as "class" or "term".