

How Matchable Are Four Thousand Ontologies on the Semantic Web

Wei Hu, Jianfeng Chen, Hang Zhang, and Yuzhong Qu

State Key Laboratory for Novel Software Technology, Nanjing University, China
{whu,yzqu}@nju.edu.cn, jf_chen@ymail.com,
hzhang.nju@gmail.com

Abstract. A growing number of ontologies have been published on the Semantic Web by various parties, to be shared for describing things. Because of the decentralized nature of the Web, there often exist different but similar ontologies from overlapped domains, or even within the same domain. In this paper, we collect more than four thousand ontologies and perform a large-scale pairwise matching based on an ontology matching tool. We create about three million mappings between the terms (classes and properties) in these ontologies, and construct a complex term mapping graph with terms as nodes and mappings as edges. We analyze the macroscopic properties of the term mapping graph as well as the derived ontology mapping graph, which characterize the global ontology matchability in several aspects, including the degree distribution, connectivity and reachability. We further establish a pay-level-domain mapping graph to understand the common interests between different ontology publishers. Additionally, we publish the generated mappings online based on the R2R mapping framework. These mappings and our observations are believed to be useful for the Linked Data community in ontology creation, integration and maintenance.

1 Introduction

The *Semantic Web* is an ongoing effort by the W3C Semantic Web Activity for realizing data integration and sharing across different applications and parties. As of today, a growing number of popular *ontologies* have emerged to describe things for specific domains, e.g., the Friend of a Friend (FOAF). These ontologies recommend common classes and properties (uniformly called *terms* in this paper) that are widely and consistently used in data sources.

Because of the decentralized nature of the Web, there usually exist multiple ontologies from overlapped application domains or even within the same domain. In order to establish interoperability between (Semantic) Web applications that use different but related ontologies, *ontology matching* (OM) has been proposed as an effective way for handling the semantic heterogeneity problem. It is useful for many tasks, such as data integration and distributed query processing.

To date, a large amount of (semi-)automatic OM approaches have been proposed in literature [10], which exploit a wide range of characteristics in ontologies,