A Formal Foundation for Ontology-Alignment Interaction Models

Marco Schorlemmer, IIIA, CSIC, Spain
Yannis Kalfoglou, University of Southampton, UK
Manuel Atencia, IIIA, CSIC, Spain

ABSTRACT

Ontology alignment foundations are hard to find in the literature. The abstract nature of the topic and the diverse means of practice make it difficult to capture it in a universal formal foundation. We argue that such a lack of formality hinders further development and convergence of practices, and in particular, prevents us from achieving greater levels of automation. In this article we present a formal foundation for ontology alignment that is based on interaction models between heterogeneous agents on the Semantic Web. We use the mathematical notion of information flow in a distributed system to ground our three hypotheses of enabling semantic interoperability and we use a motivating example throughout the article: how to progressively align two ontologies of research quality assessment through meaning coordination. We conclude the article with the presentation—in an executable specification language—of such an ontology-alignment interaction model.

Keywords: agent interaction; information-flow theory; ontologies; semantic matching

INTRODUCTION

Semantic heterogeneity is a phenomenon that emerges mostly in distributed heterogeneous environments, and it is addressed by a wide variety of communities and through the application of many diverse technologies. Its roots date back to the early stages of federated databases (Sheth & Larson, 1990) and has been continuously under investigation by database researchers through the application of a variety of (semi-)automatic schema matching techniques like those listed in Rahm and Bernstein (2001). It is well known that, for two separate systems to be capable of interoperating, exchanging vocabulary and syntax is insufficient because
one needs also to agree upon the meaning of the communicated syntactic constructs. Separate systems, though, are most often engineered assuming different, sometimes even incompatible, conceptualizations. Ontologies have been advocated as a solution to this semantic heterogeneity: separate systems would need to match their own conceptualizations against a common ontology of the application domain, so that all communication is done according to the constraints derived from the ontology.

Although the use of ontologies may indeed favor semantic interoperability, it relies on the existence of agreed domain ontologies in the first place. Furthermore, these ontologies will have to be as complete and as stable for a domain as possible, because different versions only introduce more semantic heterogeneity. The use of ontologies for semantic integration is more in tune with a classical codification-centred knowledge management tradition, as put forward by Corrêa da Silva and Agustí (2003). Such tradition comprises the efforts to define standard upper-level ontologies such as CyC (Lenat, 1995) and SUO (Standard Upper Ontology Working Group, 2003), or to establish public ontology repositories for specific domains to favor knowledge reuse such as the Ontolingua server (Farquhar, Fikes, & Rice, 1997). Corrêa da Silva and Agustí (2003) remark that “centralised ontologies... promise to bring the control of the organization back to what was possible under classical management techniques. The problem is that they may also bring back the rigidity of agencies organized under the classical management tenets” (p. 130).

Thus, semantic-integration approaches based on a priori common domain ontologies may be useful for clearly delimited and stable domains, but they are untenable and even undesirable in highly distributed, open, and dynamic environments such as the Semantic Web.

As a result, when ontology engineers begun to apply their products to the Semantic Web with the aim of solving the semantic heterogeneity problem, it became apparent that it would yield a new form of heterogeneity: that of ontology heterogeneity. The problem currently attracts the attention of practitioners with different backgrounds and perspectives, ranging from the database community (Doan & Halevy, 2005) to researchers investigating ontology mapping and matching approaches (Kalfoglou & Schorlemmer, 2003b; Noy, 2004; Shvaiko & Euzenat, 2005). Despite the plethora of potential solutions, however, there are issues that still remain unclear and for which the researchers do not share a universal understanding. Part of the problem is due to the fact that the majority of work in ontology mapping or database schema matching is based on techniques that use syntactic and structural features of ontologies. The emphasis is on automation, scalability and (re-)use of alignment algorithms but there is an apparent lack of formal foundations for most of this work (with the notable exceptions of Alagic and Bernstein (2002) and Bench-Capon, Malcolm, and Shave (2003), for example). Even when formal foundations and theory take front stage, like in Kent (2005), there is a lack of practical implementations that provide insight to the application of the proposed theory.

The quest is to find the right balance between theory and practice on one hand, and to bridge the gap between syntax-based and semantic-based solutions on the other. It seems that these core topics are antithetical: the more practical an approach is, the less semantically rich it is; the more syntax-oriented an approach is, the less practical it will be—arguably, since semantically rich approaches are the Holy Grail of semantic integration. This illustration of the problem is probably not applicable to all situations. Notable exceptions like Giunchiglia and Shvaiko (2003)—with a rich semantic flavor—claim to be practical, but they still represent only a fraction of reported systems (see, for example the overviews in Kalfoglou & Schorlemmer, 2003b; Noy, 2004; Shvaiko & Euzenat, 2005). Still, as we strive for semantically rich and practical approaches we need to find the right theoretical foundation that will support them.

The angle from which we approach this is by looking at the mathematics of information flow underlying semantic alignment. More
Related Content

ACRONYM: Context Metrics for Linking People to User-Generated Media Content

Enabling RDF Stream Processing for Sensor Data Management in the Environmental Domain

Community-Driven Linked Data Authoring and Production of Consolidated Linked Data

Web Data Quality: Current State and New Challenges

A Novel Framework for Image Categorization and Automatic Annotation