Ontologies, Data and Constraint Satisfaction

Carsten Lutz

Universität Bremen, Germany
`clu@uni-bremen.de`

Session: 6. Computational Logic

Ontology-Based Data Access (OBDA) is a paradigm for data access and management which has recently become popular for dealing with data sets that are incomplete, unstructured and heterogeneous in format (such as web data). The central idea of OBDA is to combine the actual data with an ontology that assigns a meaning to the data, enriches it with domain knowledge and provides the user with a uniform query vocabulary. Ontologies are typically formulated in logical languages such as description logics, which have been standardized by the W3C committee as the web ontology language OWL and can be viewed as a form of modal logic.

In this talk, I present and discuss a close connection between OBDA, constraint satisfaction problems (CSPs) and related formalisms such as the existential second-order logic MMSNP, introduced by Feder and Vardi in an attempt to find a counterpart of CSP in descriptive complexity theory. It turns out that for well-known ontology and query languages, every ontology-mediated query $(\mathcal{O}, q)$, which consists of an ontology $\mathcal{O}$ and an actual query q, is (in a natural sense) equivalent to a CSP, and vice versa. When varying the ontology and query language, analogous equivalences arise with MMSNP and interesting varitions thereof such as GMSNP, in which monadic second-order quantification is replaced with guarded second-order quantification. As a by-product, these equivalences also clarify the relative expressive power of ontology-mediates queries and more traditional database query languages, in particular disjunctive datalog.

The aim of the talk is to survey the general picture that emerges from these observations and to present results that can be obtained by taking advantage of the established equivalences. For example, classifying every ontology-mediated query $(\mathcal{O}, q)$ regarding its data complexity corresponds, for many ontology and query languages, to a full classification of CSPs regarding their complexity and thus involves solving the well-known Feder-Vardi conjecture. Recent advances in CSP allow to establish the decidability and computational complexity (sometimes) of highly relevant OBDA problems such as containment of ontology-mediated queries and rewritability of such queries into first-order queries or datalog programs.