The Bag Semantics of Ontology-Based Data Access (Extended Abstract)*

Charalampos Nikolaou, Egor V. Kostylev, George Konstantinidis, Mark Kaminski, Bernardo Cuenca Grau, and Ian Horrocks

Department of Computer Science, University of Oxford, UK

Ontology-based data access (OBDA) is an increasingly popular approach to enable uniform access to multiple data sources with diverging schemas. In OBDA, an ontology provides a unifying conceptual model for the data sources together with domain knowledge. The ontology is linked to each source by global-as-view (GAV) mappings, which assign views over the data to ontology predicates. Users access the data by means of queries formulated using the vocabulary of the ontology; query answering amounts to computing the certain answers of the query over the union of ontology and the materialisation of the views defined by the mappings. The formalism of choice for representing ontologies in OBDA is the description logic *DL-Lite*$_{\mathcal{R}}$ which underpins OWL 2 QL. *DL-Lite*$_{\mathcal{R}}$ was designed to ensure that queries against the ontology are *first-order rewritable*; that is, they can be reformulated as a set of relational queries over the sources.

An important observation about the conventional semantics of OBDA is that it is set-based: the materialisation of the views defined by the mappings is formalised as a *virtual ABox* consisting of a set of facts over the ontology predicates. This treatment is, however, in disagreement with the semantics of database views, which is based on bags (multisets) and where duplicate tuples are retained by default. The distinction between set and bag semantics in databases is very important in practice since it influences the evaluation of aggregate queries that combine various aggregation functions (e.g., Min, Max, Sum, Count, Avg) with the grouping functionality provided in SQL by the GroupBy construct.

In this paper we study a bag semantics for OBDA that is compatible with the semantics of standard databases and can provide the foundations for the future study of aggregate queries. We propose the ontology language *DL-Lite*$_{\mathcal{R}}^{bag}$ and its restriction *DL-Lite*$_{core}^{bag}$, where ABoxes consist of a bag of facts. We define the bag semantics of conjunctive query (CQ) answering and show that it is compatible with the conventional set semantics. However, we show that *DL-Lite*$_{core}^{bag}$ ontologies may not admit a universal model while CQ answering becomes coNP-hard in data complexity, hence losing first-order rewritability of CQs. To regain tractability, we study the class of *rooted CQs*, which captures most practical OBDA queries, and show that it admits universal models and enjoys first-order rewritability for *DL-Lite*$_{core}^{bag}$ ontologies. Unfortunately, these properties do not extend to *DL-Lite*$_{\mathcal{R}}^{bag}$ where CQ answering remains coNP-hard.

* This is the extended abstract of the paper accepted for publication in the proceedings of IJCAI 2017. The full paper may be accessed at `arXiv:1705.07105 [cs.AI]`. Work supported by the Royal Society under a University Research Fellowship, the EPSRC projects ED3 and DBOnto, and the Research Council of Norway via the Sirius SFI.