

Model-theoretic Characterizations of Rule-based Ontologies^{*}

Marco Console¹, Phokion G. Kolaitis², and Andreas Pieris³

¹ Sapienza, University of Rome console@diag.uniroma1.it

² UC Santa Cruz & IBM Research kolaitis@ucsc.edu

³ University of Edinburgh apieris@inf.ed.ac.uk

Model theory is the study of the interaction between formulas in some logical formalism and their models, that is, structures that satisfy the formulas. There are two directions in this interaction, namely, from syntax to semantics and from semantics to syntax. The first direction aims to identify structural properties possessed by all models of formulas having common syntactic features. For example, it is easy to show that every universal first-order sentence is preserved under substructures. The second direction aims to characterize formulas in terms of their structural properties. For example, the well known Łos-Tarski Theorem asserts that if a first-order sentence is preserved under substructures, then it is logically equivalent to a universal first-order sentence. In general, establishing results in the second direction is a much harder task than establishing results in the first. In other words, obtaining model-theoretic characterizations of formulas or of classes of formulas is a far greater challenge than identifying structural properties possessed by all models of formulas with common syntactic features.

Makowsky and Vardi [9] were the first to obtain model-theoretic characterizations of classes of database dependencies expressed in suitable fragments of first-order logic. Furthermore, they classified the work on model-theoretic characterizations into two distinct approaches, which they called the *preservation* and the *axiomatizability* approach:

Preservation Approach. One considers two logical formalisms \mathcal{L} and \mathcal{L}' , where \mathcal{L}' is typically a proper fragment of \mathcal{L} , and the goal is to obtain model-theoretic characterizations of the following form: a set Σ of \mathcal{L} -formulas is equivalent to a set Σ' of \mathcal{L}' -formulas if and only if the models of Σ satisfy certain structural properties. For example, the Łos-Tarski Theorem is a prime example in the preservation approach.

Axiomatizability Approach. One considers a logical formalism \mathcal{L} , and the goal is to obtain model-theoretic characterizations of the following form: a class \mathcal{C} of structures is axiomatizable by a (possibly infinite) set of \mathcal{L} -formulas if and only if the structures in \mathcal{C} satisfy certain structural properties. Makowsky and Vardi [9] further distinguished the special case of *finite axiomatizability*, where the goal is to obtain model-theoretic characterizations of when a class of structures is axiomatizable by a *finite* set of \mathcal{L} -formulas. They

^{*} This is an extended abstract of [4]. “Copyright © 2021 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).”

then obtained model-theoretic characterizations of axiomatizability and finite axiomatizability of classes of database dependencies.

Database dependencies were originally used to formalize integrity constraints in databases with much of the early work in this area focusing on the implication problem between database dependencies (see [6] for a survey). A prominent class of database dependencies is that of tuple-generating dependencies (tgds), that is, first-order sentences of the form $\forall \bar{x} \forall \bar{y} (\phi(\bar{x}, \bar{y}) \rightarrow \exists \bar{z} \psi(\bar{x}, \bar{z}))$, where $\phi(\bar{x}, \bar{y})$ is a (possibly empty) conjunction of atoms, and $\psi(\bar{x}, \bar{z})$ is a (non-empty) conjunction of atoms. Later on, tgds found numerous uses in other data management tasks. They have been used as schema-mapping specification languages and, as such, have been successfully deployed in the study of data exchange [1, 5] and data integration [7]. More recently, tgds have been used for modeling ontologies. In fact, tgds have been extensively studied in the context of ontologies under different names, such as Datalog+/- [2] and existential rules [10].

Several model-theoretic characterizations following the preservation approach have been established for tgds [11], as well as for description logics [8]. Concerning the axiomatizability/finite axiomatizability approach, model-theoretic characterizations of full tgds (i.e., tgds with no existentially quantified variables in the right-hand side) were obtained in [9], while such characterizations of source-to-target tgds were obtained in [3] (source-to-target tgds have been used to formalize data exchange between a source schema and a target schema). These results notwithstanding, however, the study of model-theoretic characterizations of sets of arbitrary tgds in the axiomatizability/finite axiomatizability approach has remained largely unexplored so far.

Summary of Results. Motivated by the preceding state of affairs, we embark here on a systematic investigation of model-theoretic characterizations of classes of tgds in the finite axiomatizability approach. This investigation is carried out in the context of ontologies. From a syntactic point of view, an ontology is specified by a set of formulas in some formalism. From a semantic point of view, an ontology can be identified with the set of all structures (finite or infinite) that satisfy the formulas specifying the ontology. Hence, as a semantic object, an ontology is an isomorphism-closed class of (finite or infinite) structures over some fixed relational schema. Our goal is to answer the following question: *what are necessary and sufficient conditions for an ontology (an isomorphism-closed class of structures) to be specified by a set of tgds?* The main outcome of this investigation is to characterize the ontologies that are finitely axiomatizable by a finite set of arbitrary tgds or by a finite set of tgds that belong to one of the main subclasses of tgds, namely, full, frontier-guarded, guarded, and linear tgds.

Our model-theoretic characterizations make use of structural properties encountered in earlier model-theoretic characterizations, such as the ontology being closed under direct products, and containing critical structures of every finite cardinality, where a structure is *critical* if each of its relations contains all possible tuples from the domain of the structure. The main innovation, however, is the introduction and use of the notion of (n, m) -locality, where, intuitively, n represents the number of universal quantifiers in the tgds and m represents the

number of existential quantifiers in the tgds. Our first main result asserts that an ontology O is axiomatizable by a finite set of tgds if and only if O is closed under direct products, contains critical structures of every finite cardinality, and is (n, m) -local for some non-negative integers n and m . The notion of (n, m) -locality turns out to be delicate, yet flexible. Indeed, this notion can be tailored to other classes of tgds, so that it gives rise to the refined notions of frontier-guarded (n, m) -locality, guarded (n, m) -locality, and linear (n, m) -locality. Using these refined notions, we obtain model-theoretic characterizations of ontologies that are axiomatizable by a finite set of, respectively, frontier-guarded, guarded, and linear tgds.

Acknowledgments. Part of Marco Console’s work was done while he was a research associate at the University of Edinburgh. Part of Andreas Pieris’ work was done while he was visiting the University of California, Santa Cruz. Marco Console has been partially supported by MIUR under the PRIN 2017 project “HOPE” (prot. 2017MMJJRE), and by the EU under the H2020-EU.2.1.1 project TAILOR, grant id. 952215. Phokion Kolaitis has been partially supported by NSF Award No. 1814152. Andreas Pieris was supported by the EPSRC grant EP/S003800/1 “EQUID”.

References

1. Arenas, M., Barceló, P., Libkin, L., Murlak, F.: Foundations of Data Exchange. Cambridge University Press (2014)
2. Cali, A., Gottlob, G., Lukasiewicz, T., Marnette, B., Pieris, A.: Datalog+/-: A family of logical knowledge representation and query languages for new applications. In: LICS. pp. 228–242 (2010)
3. ten Cate, B., Kolaitis, P.G.: Structural characterizations of schema-mapping languages. In: ICDT. pp. 63–72 (2009)
4. Console, M., Kolaitis, P.G., Pieris, A.: Model-theoretic characterizations of rule-based ontologies. In: PODS. pp. 416–428 (2021)
5. Fagin, R., Kolaitis, P.G., Miller, R.J., Popa, L.: Data exchange: semantics and query answering. Theor. Comput. Sci. 336(1), 89–124 (2005)
6. Fagin, R., Vardi, M.Y.: The Theory of Data Dependencies - A Survey. In: Symposia in Applied Mathematics, pp. 19–71 (1986)
7. Lenzerini, M.: Data integration: A theoretical perspective. In: PODS. pp. 233–246 (2002)
8. Lutz, C., Piro, R., Wolter, F.: Description logic tboxes: Model-theoretic characterizations and rewritability. In: IJCAI. pp. 983–988 (2011)
9. Makowsky, J.A., Vardi, M.Y.: On the expressive power of data dependencies. Acta Inf. 23(3), 231–244 (1986)
10. Mugnier, M., Thomazo, M.: An introduction to ontology-based query answering with existential rules. In: Reasoning Web. pp. 245–278 (2014)
11. Zhang, H., Zhang, Y., Jiang, G.: Model-theoretic characterizations of existential rule languages. In: IJCAI (2020)