On the Language of Nested Tuple Generating Dependencies

PHOKION G. KOLAITIS, UC Santa Cruz and IBM Research - Almaden REINHARD PICHLER, TU Wien EMANUEL SALLINGER, TU Wien and University of Oxford VADIM SAVENKOV, Vienna University of Economics and Business

During the past 15 years, schema mappings have been extensively used in formalizing and studying such critical data interoperability tasks as data exchange and data integration. Much of the work has focused on GLAV mappings, i.e., schema mappings specified by source-to-target tuple-generating dependencies (s-t tgds), and on schema mappings specified by second-order tgds (SO tgds), which constitute the closure of GLAV mappings under composition. In addition, nested GLAV mappings have also been considered, i.e., schema mappings specified by nested tgds, which have expressive power intermediate between s-t tgds and SO tgds. Even though nested GLAV mappings have been used in data exchange systems, such as IBM's Clio, no systematic investigation of this class of schema mappings has been carried out so far. In this article, we embark on such an investigation by focusing on the basic reasoning tasks, algorithmic problems, and structural properties of nested GLAV mappings. One of our main results is the decidability of the implication problem for nested tgds. We also analyze the structure of the core of universal solutions with respect to nested GLAV mappings and develop useful tools for telling apart SO tgds from nested tgds. By discovering deeper structural properties of nested GLAV mappings, we show that also the following problem is decidable: Given a nested GLAV mapping, is it logically equivalent to a GLAV mapping?

$\label{eq:CCS Concepts: \bullet Theory of computation} \rightarrow Data exchange; \bullet Information systems \rightarrow Mediators and data integration;$

Additional Key Words and Phrases: Schema mappings, data integration, data exchange, nested dependencies, second-order dependencies

© 2020 Copyright held by the owner/author(s). Publication rights licensed to ACM.

https://doi.org/10.1145/3369554

The research of Reinhard Pichler, Emanuel Sallinger, and Vadim Savenkov was supported by the Austrian Science Fund, projects (FWF):P25207-N23 and (FWF):Y698, and the Vienna Science and Technology Fund, project ICT12-015. The research of Phokion Kolaitis on this article was partially supported by NSF Grant IIS-1217869. The full version was completed while Kolaitis was visiting the Simons Institute for the Theory of Computing during the fall of 2016. The research of Emanuel Sallinger was supported by the EPSRC programme grant EP/M025268/1, the Horizon 2020 grant 809965 and the Vienna Science and Technology Fund (WWTF) WWTF grant VRG18-013.

Authors' addresses: P. G. Kolaitis, University of California, Santa Cruz, Computer Science and Engineering Department, CA 95064, USA and IBM Research - Almaden, San Jose, CA 95120, USA; email: kolaitis@ucsc.edu; R. Pichler, TU Wien, Faculty of Informatics, Institute of Logic and Computation, Database and Artificial Intelligence Group, 1040 Vienna, Austria; email: pichler@dbai.tuwien.ac.at; E. Sallinger, TU Wien, Faculty of Informatics, Institute of Logic and Computation, Database and Artificial Intelligence Group, 1040 Vienna, Austria; email: pichler@dbai.tuwien.ac.at; E. Sallinger, TU Wien, Faculty of Informatics, Institute of Logic and Computation, Database and Artificial Intelligence Group, Knowledge Graph Lab, 1040 Vienna, Austria and University of Oxford, Department of Computer Science, OX1 3QD, Oxford, United Kingdom; email: sallinger@dbai.tuwien.ac.at; V. Savenkov, Vienna University of Economics and Business, Institute for Information Business, Welthandelsplatz 1, 1020 Vienna, Austria; email: vadim.savenkov@wu.ac.at.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

^{0362-5915/2020/07-}ART8 \$15.00