Scalable Reasoning on Document Stores via Instance-Aware Query Rewriting

Olivier Rodriguez, Federico Ulliana, Marie-Laure Mugnier

▶ To cite this version:

Scalable Reasoning on Document Stores via Instance-Aware Query Rewriting

Olivier Rodriguez
LIRMM, Inria, Univ. Montpellier,
CNRS
Montpellier, France
olivier.rodriguez@inria.fr

Federico Ulliana
LIRMM, Inria, Univ. Montpellier,
CNRS
Montpellier, France
federico.ulliana@inria.fr

Marie-Laure Mugnier
LIRMM, Inria, Univ. Montpellier,
CNRS
Montpellier, France
marie-laure.mugnier@inria.fr

ABSTRACT

Data trees, typically encoded in JSON, are ubiquitous in data-driven applications. This ubiquity makes urgent the development of novel techniques for querying heterogeneous JSON data in a flexible manner. We propose a rule language for JSON, called constrained tree-rules, whose purpose is to provide a high-level unified view of heterogeneous JSON data and infer implicit information. As reasoning with constrained tree-rules is undecidable, we identify a relevant subset featuring tractable query answering, for which we design an automata-based query rewriting algorithm. Our approach consists of leveraging NoSQL document stores by means of a novel instance-aware query-rewriting technique. We present an extensive experimental analysis on large collections of several million JSON records. Our results show the importance of instance-aware rewriting as well as the efficiency and scalability of our approach.