

The Query Containment Problem: Set Semantics vs. Bag Semantics

Phokion G. Kolaitis

UC Santa Cruz and IBM Research - Almaden

Abstract. Query containment is a fundamental algorithmic task in database query processing and optimization. Under set semantics, the query-containment problem for conjunctive queries has long been known to be NP-complete. SQL queries, however, are typically evaluated under bag semantics and return multisets as answers, since duplicates are not eliminated unless explicitly specified. The exact complexity of the query-containment problem for conjunctive queries under bag semantics has been an outstanding and rather poorly understood open problem for twenty years. In fact, to this date, it is not even known whether conjunctive-query containment under bag semantics is decidable. The goal of this talk is to draw attention to this fascinating problem by presenting a comprehensive overview of old and not-so-old results about the complexity of the query-containment problem for conjunctive queries and their variants, under both set semantics and bag semantics.