

Computing Range Consistent Answers to Aggregation Queries via Rewriting

Aziz Amezian El Khalfioui and Jef Wijsen

University of Mons, Belgium
{aziz.amezianelkhalfioui,jef.wijsen}@umons.ac.be

We consider the problem of answering conjunctive queries with aggregation on database instances that may violate primary key constraints. In SQL, these queries follow the SELECT-FROM-WHERE-GROUP BY format, where the WHERE-clause involves a conjunction of equalities, and the SELECT-clause can incorporate aggregate operators like MAX, MIN, SUM, AVG, or COUNT. Repairs of a database instance are defined as inclusion-maximal subsets that satisfy all primary keys. For a given query, our primary objective is to identify repairs that yield the lowest aggregated value among all possible repairs. We particularly investigate queries for which this lowest aggregated value can be determined through a rewriting in first-order logic with aggregate operators.

This problem is particularly related to the Consistent Query Answering problem, which is the problem that asks whether a boolean conjunctive query is verified by every repair of a database instance that may violate its constraints. In this work, we use tools and properties developed in the context of the Consistent Query Answering problem and extend them to study the problem that is the focus of this work.

This paper was accepted during the first submission phase of ACM PODS 2025 [KW24b], which will take place in June 2025 in Berlin, Germany. The arXiv version is [KW24a].

References

- [KW24a] Aziz Amezian El Khalfioui and Jef Wijsen. Computing range consistent answers to aggregation queries via rewriting. *CoRR*, abs/2409.01648, 2024.
- [KW24b] Aziz Amezian El Khalfioui and Jef Wijsen. Computing range consistent answers to aggregation queries via rewriting. *Proc. ACM Manag. Data*, 2(2), To be published in November 2024.