

# Automatic Discovery and Verification of Database Query Transformations



## Challenge:

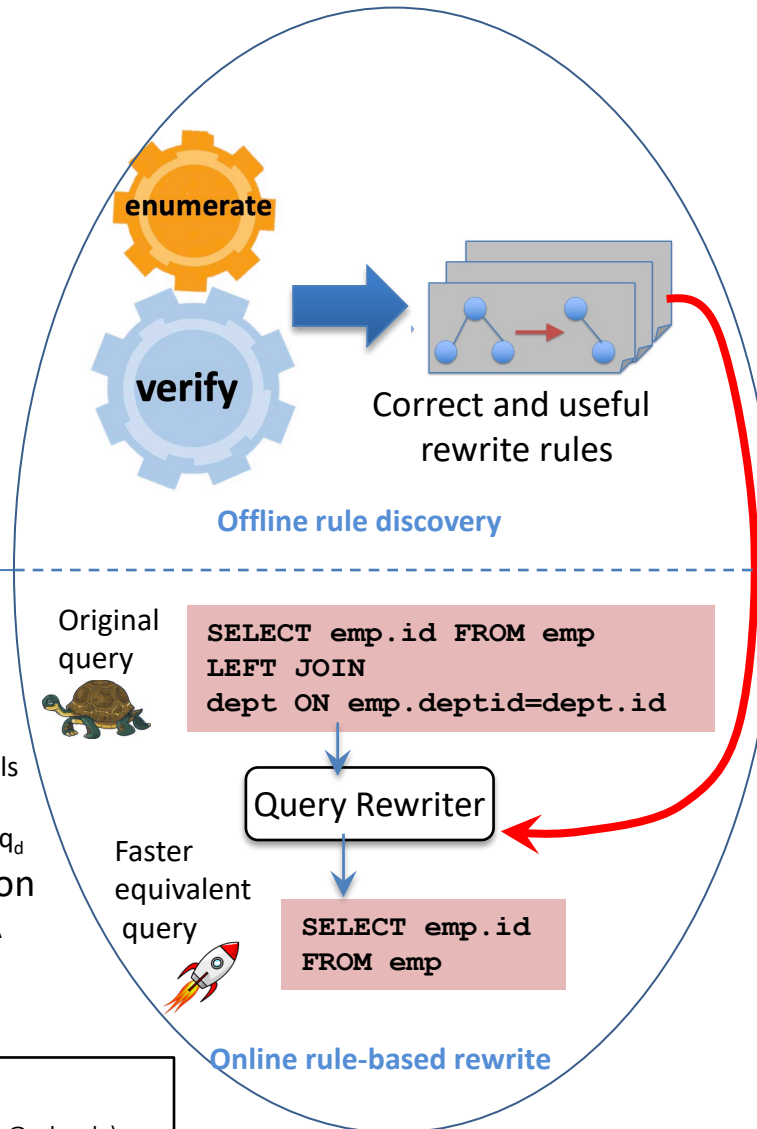
- Rule Discovery:
  - How to define the search space and enumerate rules?
- Rule Verification:
  - How to prove query equivalence based on semantics instead of syntax?

## Scientific Impact:

- The first tool for the automatic discovery of SQL rewrite rules.
- State-of-the-art SQL equivalence checker based on the theory of LIA\*
- A new theory extending general SMT theory with multisets and Presburger arithmetic

## Solution:

- Model rules as  $\langle q_s, q_d, c \rangle$  for enumeration
  - $q_s, q_d$ : templated queries with symbols for table/column names.
  - $c$ : a constraint relating symbols of  $q_s, q_d$
- Semantics-based rule verification
  - Query  $\rightarrow$  U-expression  $\rightarrow$  LIA\*  $\rightarrow$  LIA
  - LIA\* enables reasoning “sums with unbounded domain” in U-expr



## Broader Impact and Broader Participation:

- All database-backed applications benefit from faster queries.
  - Newly discovered rules have fixed 38 out of 50 Github issues in Gitlab etc.
- Our verifier is being used by database classes for auto-grading
- Provided undergrad research opportunity
- Created new course materials at NYU

Award ID: fmitf-2220407

New York University, Yale University

Jinyang Li (jinyang@cs.nyu.edu), Ruzica Piskac (ruzica.piskac@yale.edu)



The NSF Formal Methods in the Field PI Meeting (2024 FMitF PI Meeting)  
November 12-13, 2024 | The University of Iowa | Iowa City, Iowa