ProvSQL: Provenance and Probability Management in PostgreSQL

Pierre Senellart, Louis Jachiet, Silviu Maniu, Yann Ramusat

ProvSQL in Brief

- Light-weight extension for PostgreSQL ≥ 9.5
- Transparent computation of provenance of SQL queries as a term algebra circuit
- Allows probabilistic query evaluation with various techniques:
  - Enumeration of possible worlds
  - Knowledge compilation to d-DNNF via external tools (C2d, Dsharp, D4)
  - Monte-Carlo sampling
- Open-source and available at https://github.com/PierreSenellart/provsql

Supported Queries and Provenance Annotations

- Large subset of (nested) SQL, without aggregates or UDFs:
  - SELECT ... FROM ... WHERE, JOIN, DISTINCT, GROUP BY
  - UNION, UNION ALL, EXCEPT
- Variety of provenance formalisms:
  - arbitrary provenance semirings
  - arbitrary provenance semirings with monus (m-semirings)
  - where-provenance

How ProvSQL Computes Various Provenance Forms and Probabilities

The database with provenance annotations is used to compute provenance forms and probabilities. The query evaluation process involves selecting distinct cities from the Personnel table except for those in P1 and P2, with additional conditions on positions. The circuit in the universal m-semiring is used to count m-semiring terms, specifically for bag semantics.

References