QOCO-R:Query and Rules Oriented Data Cleaning

ABSTRACT

Fill Content

1. INTRODUCTION

Fill intro

2. RULES

The rules in QOCO-Rre a database assertions that delivered to the algorithm by the Oracle. In order to get an effective rules language it must enable expressing all popular relational database assertions. As in [1, 2] The embedded dependencies include all of the naturally-occurring constraints on relational databases. They are a first order logic formulas of the form:

$$\forall \mathtt{x}_1, ..., \mathtt{x}_n \phi(\mathtt{x}_1, ..., \mathtt{x}_n) \rightarrow \exists \mathtt{z}_1, ..., \mathtt{z}_m \psi(\mathtt{x}_1, ..., \mathtt{x}_n, \mathtt{z}_1, ..., \mathtt{z}_m)$$

where the left hand side (LHS) of the implication, ϕ, is a conjunction of relational formulas over variables $\bar{x}$, and the right hand side (RHS) of the implication, ψ, is a conjunction of relational or equality formulas over variables $\bar{x}$ and $\bar{z}$. The embedded dependencies is comprised of tuple-generating dependencies (tgds) of the form:

$$\forall \mathtt{x}_1, ..., \mathtt{x}_n \phi(\mathtt{x}_1, ..., \mathtt{x}_n) \rightarrow \exists \mathtt{z}_1, ..., \mathtt{z}_m \mathtt{R}(\mathtt{x}_1, ..., \mathtt{x}_n, \mathtt{z}_1, ..., \mathtt{z}_m)$$

and equality generating dependencies (egds) of the form:

$$\forall \mathtt{x}_1, ..., \mathtt{x}_n \phi(\mathtt{x}_1, ..., \mathtt{x}_n) \rightarrow \mathtt{x}_i = \mathtt{x}_j$$

In tgds the RHS contains only relational formulas and in egds the RHS contains only equality formulas. Given a particular combination of tuples satisfying the constraint of the LHS, tgds expresses an assertion about the existence of a tuple in the instance on the RHS,and egds expresses an assertion about an equality between two variables.
As we mentioned above our rules language should enable expressing tgds and egds, therefor the rules language consists of two sets of rules:

1. Tuple-generating rules (tgrs). They have the same form as tgds but the LHS may contain also constraints on variables (not only relational formulas), a constraint is a boolean expression of the form $v\,OP\,w$ where $v, w \in \bar{x} \cup \mathcal{C}$ and OP is a boolean operation that defined on the variables domain, for example if v and w value should be a number then OP should be $=, \neq, \leq, \geq, ...$ For
2. Condition-generating rules (cgrs). They have the same form as egds but both the LHS and the RHS could contain a conjunction of constraints.

3. CONCLUSIONS

Content

4. ACKNOWLEDGMENTS

Content

5. REFERENCES

[1] A. Deutsch, L.Popa, and V. Tannen. *Query reformulation with constraints. SIGMOD Record, 35(1), 2006.*

[2] R. Fagin, P. Kolaitis, R. J. Miller, , and L. Popa. *Data exchange: Semantics and query answering. Theoretical Computer Science, 336, 2005.*

APPENDIX

Content