A framework for Testing Information Preservation in Schema Transformations

Nonyelum Ndefo
KRDB Research Centre for Knowledge and Data
Dominikanerplatz 3 - Piazza Domenicani
Bozen-Bolzano, Italy
nndefo@unibz.it

ABSTRACT

In this work we offer a constructive framework to test information preservation between two first-order schemas guided by model-theoretic concepts. The presented work complements existing definitions by presenting specific details on the mechanism involved in testing for information capacity preservation.

KEYWORDS

Schema Transformation, Schema Dominance, Database Design

ACM Reference Format:
Nonyelum Ndefo. 2018. A framework for Testing Information Preservation in Schema Transformations. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 1 page. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

According to established literature, a schema transformation occurs between two schemas when the instances of a source schema are transformed into instances of a target schema via well defined mapping functions. Following any transformation, it is expected that the new schema contains at least the same information as the original. This is called *information capacity preservation* or *schema dominance*. A target schema dominates a source schema when all the instances of the source schema are preserved in the target. Both schemas are equivalent if they dominate each other. We call this characteristic, losslessness. Despite the literature, there has been no detailed systematic approach to proving that a schema truly dominates another one.

2 RELATED WORK

To compare the information capacity between relational schemas, Hull [2] defines four progressively restrictive notions of schema dominance, the most restrictive being a type of dominance involving expressions in first order predicate calculus. These results are further developed by [1] in application to relational schemas with key dependencies. Miller et al. in [3] place emphasis on the relationship between schemas within the context of schema integration.

Woodstock '18, June 03–05, 2018, Woodstock, NY

ACM ISBN 978-1-4503-9999-9/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

3 METHODOLOGY

The main idea of our work is to reduce the detection of schema dominance and losslessness to a problem of entailment checking involving the set of rules (constraints) which restrict the schemas, and a pair of structural mappings called First-Order (FO) mappings: a set of views between the schemas' diverse alphabets which describe how each structure in one schema is defined as a query over the other. These FO mappings are the bases for the functions which transform the valid instances of one schema into instances of the other. We perceive the test for schema dominance as a question: *Do the constraints of the source schema and the FO mapping defining the structures in the target schema logically entail the constraints of the target schema and an inverse FO mapping?* For losslessness, we do a bidirectional test.

Since we are dealing with schemas expressed in first order predicate logic, we take advantage of an automated theorem prover (Prover9) to automate the entailment checking on concrete transformation scenarios. The tool either verifies the satisfiability of the entailment, or searches for possible counter examples

The main contribution involves : 1) detailing, through examples, the connection between the models of the premise(s) and of the conclusion of the entailment in the case of schema dominance; 2) verifying such entailment to be equivalent to the existing notion of schema dominance; 3) reviewing, through our framework, the lossless join and dependency preservation properties found in database design theory; and the requirement for schema equivalence in the database reverse engineering process.

We discover that our framework can be used as generalized verification technique for database normalization, and for verifying the accuracy of database reverse engineering transformations. We expect that adopting such a framework will be beneficial to database engineers in contributing to richer schema transformation methodologies in fields such as schema integration, database decompositions, and database reverse engineering.

REFERENCES

[1] J. Albert, Y. Ioannidis, and R. Ramakrishnan. Equivalence of keyed relational schemas by conjunctive queries. *Journal of Computer and System Sciences*, 58(3):512–534, 1999.
[2] R. Hull. Relative information capacity of simple relational database schemata. *SIAM Journal on Computing*, 15(3):856–886, 1986.
[3] R. J. Miller, Y. E. Ioannidis, and R. Ramakrishnan. Schema equivalence in heterogeneous systems: bridging theory and practice. *Information Systems*, 19(1):3–31, 1994.