TU Clausthal

Kolloquiums-Reihe des Instituts für Informatik

Mittwoch, 10.05.2023, 15 Uhr c.t.
Institut für Informatik, Julius-Albert-Str. 4, Raum 106
und BBB https://webconf.tu-clausthal.de/b/ger-9in-yfv-ifn

Prof. Dr. Joachim Biskup

Probabilistic Instance Generation, Functional Constraints, Relational Schema Normalization, and Counting Properties

Abstract

For the experimental evaluation of software intended to operate on relation instances for a given relational schema with functional constraints and size restrictions on the finite attribute domains, we studied the problem of uniform probabilistic generation of sample instances of predetermined size. We completely solved the problem for schemas with just one functional dependency: while a non-normalized schema requires a sophisticated procedure of high computational complexity based on the integer partition problem, a schema in Object Normal Form (Boyce-Codd Normal Form and Unique Key) allows a straightforward generation heuristic. Moreover, in the normalized case, sampling of representatives of classes under equivalence by bijective transformations and by combinatorial similarity, respectively, coincide and, based on that, the generation heuristic is also successful for sampling with a probability distribution that is uniform regarding the size of the classes.

Besides solving the original problem, our studies provided further insight to the topic of functional dependencies when supplemented with domain-size restrictions. Furthermore, achieving a uniform probability distribution turned out to be closely related to schema normalization. Finally, fixing the size of relation instances to be generated leads to the involvement of integer partitions and, thus, of number theory, despite the lack of an obvious "real-world" application meaning. A tentative abstraction from that setting results in the challenge of dealing with both multiplicity bounds as a special cardinality constraint on elementary selections and diversity bounds on projections on attribute sets.