TECHNISCHE UNIVERSITÄT DRESDEN

Faculty of Computer Science Institute of Theoretical Computer Science, Chair of Automata Theory

Database Theory

Summer Semester 2016

Exercise Sheet 9 – Datalog

20th June 2016

Dr. rer. pol. Markus Krötzsch & Dipl.-Math. Francesco Kriegel

Exercise 9.1 Show that any Datalog program can be expressed as a safe Datalog program that is polynomial in size of the original program and given schema.

Exercise 9.2 Assume that the database uses a binary EDB predicate edge to store a directed graph. Try to express the following properties in semipositive Datalog programs with a successor ordering, or explain why this is not possible.

(a) The database contains an even number of elements.

(b) The graph contains a node with two outgoing edges.

(c) The graph is 3-colourable.

(d) The graph is *not* connected (*).

(e) The graph does not contain a node with two outgoing edges.

(f) The graph is a chain.

Exercise 9.3 In the lecture, we have used a restricted form of propositional Horn logic where all rules are of the form $H \leftarrow$ or $H \leftarrow B_1 \wedge B_2$. We refer to this logic as *propHorn2*.

It was claimed that entailment checking in *propHorn2* is P-hard. To support this claim, explain how entailment in propositional Horn logic can be reduced to entailment in *propHorn2*. Argue how this reduction can be accomplished in logarithmic space.

Exercise 9.4 Prove that entailment checking in propositional Horn logic is P-hard.

Hint. Modify the ExpTime Turing machine simulation from the lecture to simulate a PTime Turing machine instead.

Exercise 9.5 Show that the following property cannot be expressed in Datalog:

> The edge predicate has a *proper* cycle, i.e., a cycle that is not of the form edge(a, a).

Can you express this property using ...

(a) ... a successor ordering?

(b) ... atomic negation?

(c) ... an equality predicate $\approx$ with the obvious semantics?

(d) ... an inequality predicate $\not\approx$ with the obvious semantics?

Exercise 9.6 Consider the query mapping that returns all pairs of elements in a database that are connected by a directed edge path of length ℓ^2 for some $\ell \geq 0$. Paths do not have to be simple, i.e., the same edge might be used more than once.

(a) Show that this query mapping is closed under homomorphisms.

(b) Show that this query cannot be expressed in Datalog $(*)$.

Hint. for (b), show a kind of "pumping lemma". In detail, show that, for every Datalog program P, there is a number $\ell \geq 0$, such that every proof tree that entails an answer for a path of length ℓ^2 can be extended ("pumped") to obtain larger proof trees that recognise paths of lengths that are not square numbers.