

Exercise Sheet 11: Graph Databases and Path Queries

Maximilian Marx, Markus Krötzsch

Database Theory, 2022-06-28, Summer Term 2022

Exercise 11.1. It was explained in the lecture that RDF and Property Graph can encode the same graph structures. How could we encode arbitrary hypergraphs (relational databases) in RDF? RDF can be considered as a synonym for “labelled directed graph” here – the technical details of the RDF standard are not important for this exercise.

Exercise 11.2. Can the following Datalog programs be encoded using a C2RPQ? In each case, give a suitable C2RPQ or explain why there is none.

1. The “Same generation” Datalog program from the lecture:

$$\begin{aligned} S(x, x) &\leftarrow \text{human}(x) \\ S(x, y) &\leftarrow \text{parent}(x, w) \wedge S(w, w) \wedge \text{parent}(y, v) \end{aligned}$$

2. Ancestors born in the same city:

$$\begin{aligned} \text{AncCity}(x, y, x', y') &\leftarrow \text{parent}(x, x') \wedge \text{bornIn}(x, y) \wedge \text{bornIn}(x', y') \\ \text{AncCity}(x, y, x', y') &\leftarrow \text{AncCity}(x, y, w, v) \wedge \text{AncCity}(w, u, x', y') \\ \text{Query}(x, x', y) &\leftarrow \text{AncCity}(x, y, x', y) \end{aligned}$$

3. Ancestors of Dresden-based family lines:

$$\begin{aligned} \text{DDAnc}(x, y) &\leftarrow \text{parent}(x, y) \wedge \text{bornIn}(x, \text{dresden}) \wedge \text{bornIn}(y, \text{dresden}) \\ \text{DDAnc}(x, z) &\leftarrow \text{DDAnc}(x, y) \wedge \text{parent}(y, z) \wedge \text{bornIn}(z, \text{dresden}) \end{aligned}$$

Exercise 11.3. Consider the method for checking RPQ containment as sketched on slide “Containment for RPQs” in the lecture. Explain the procedure and the resulting complexity bounds in your own words. How could one construct the required automaton “on the fly”?

Exercise 11.4. Give an example for a binary C2RPQ that cannot be expressed as a 2RPQ. By a *linear binary C2RPQ* we mean a C2RPQ of the form

$$\exists x_{k_1}, \dots, x_{k_m}. R_1(x_1, x_2) \wedge R_2(x_2, x_3) \wedge \dots \wedge R_{n-1}(x_{n-1}, x_n)$$

where each $R_i(x_i, x_{i+1})$ is an atom or a 2RPQ, and the x_{k_j} are among the variables that occur in the query. Can every linear binary C2RPQ be expressed by a 2RPQ? Explain your answer.

Exercise 11.5. Give an example of a Datalog query that contains both of the following (and maybe also other) rules

$$\begin{aligned} \text{Query}(x, z) &\leftarrow p_a(x, y) \wedge p_b(y, z) \\ \text{Query}(x, z) &\leftarrow p_a(x, x') \wedge \text{Query}(x', z') \wedge p_b(z', z) \end{aligned}$$

and that can be expressed as a C2RPQ.