Foundations of Databases and Query Languages
Exercise 12: Graph Databases and Path Queries
17 July 2015

Exercise 12.1 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 12.2 Can the following Datalog programs be encoded using a C2RPQ? In each case, give a suitable C2RPQ or explain why there is none.

(a) The "Same generation" Datalog program from the previous exercise:

$$\mathsf{S}(x, x) \leftarrow \mathsf{human}(x)$$
$$\mathsf{S}(x, y) \leftarrow \mathsf{parent}(x, w) \wedge \mathsf{S}(v, w) \wedge \mathsf{parent}(y, v)$$

(b) Ancestors born in the same city:

$$\mathsf{AncCity}(x, y, x', y') \leftarrow \mathsf{parent}(x, x') \wedge \mathsf{bornIn}(x, y) \wedge \mathsf{bornIn}(x', y')$$
$$\mathsf{AncCity}(x, y, x'', y'') \leftarrow \mathsf{AncCity}(x, y, x', y') \wedge \mathsf{AncCity}(x', y', x'', y'')$$
$$\mathsf{Query}(x, x', y) \leftarrow \mathsf{AncCity}(x, y, x', y)$$

(c) Ancestors of Dresden-based family lines:

$$\mathsf{DDAnc}(x, y) \leftarrow \mathsf{parent}(x, y) \wedge \mathsf{bornIn}(x, \mathsf{dresden}) \wedge \mathsf{bornIn}(y, \mathsf{dresden})$$
$$\mathsf{DDAnc}(x, z) \leftarrow \mathsf{DDAnc}(x, y) \wedge \mathsf{parent}(y, z) \wedge \mathsf{bornIn}(z, \mathsf{dresden})$$

Exercise 12.3 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 12.4 Give an example for a binary C2RPQ that cannot be expressed as a 2RPQ.
  By a *linear C2RPQ* we mean a C2RPQ of the form

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

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 binary linear C2RPQ be expressed by a 2RPQ? Explain your answer.

Exercise 12.5 Give an example of a Datalog query that contains both of the following (and maybe also other) rules

$$\mathsf{Query}(x, z) \leftarrow \mathsf{p}_a(x, y) \wedge \mathsf{p}_b(y, z)$$
$$\mathsf{Query}(x, z) \leftarrow \mathsf{p}_a(x, x') \wedge \mathsf{Query}(x', z') \wedge \mathsf{p}_b(z', z)$$

and that can be expressed as a C2RPQ.