

**Exercise Sheet 8: Datalog**  
 Maximilian Marx, Markus Krötzsch  
 Database Theory, 2022-05-31, Summer Term 2022

**Exercise 8.1.** Consider the example Datalog program from the lecture:

father(alice, bob)	(0.1)
mother(alice, carla)	(0.2)
mother(ewan, carla)	(0.3)
father(carla, david)	(0.4)
Parent( $x, y$ ) $\leftarrow$ father( $x, y$ )	(0.5)
Parent( $x, y$ ) $\leftarrow$ mother( $x, y$ )	(0.6)
Ancestor( $x, y$ ) $\leftarrow$ Parent( $x, y$ )	(0.7)
Ancestor( $x, z$ ) $\leftarrow$ Parent( $x, y$ ) $\wedge$ Ancestor( $y, z$ )	(0.8)
SameGeneration( $x, x$ )	(0.9)
SameGeneration( $x, y$ ) $\leftarrow$ Parent( $x, v$ ) $\wedge$ Parent( $y, w$ ) $\wedge$ SameGeneration( $v, w$ )	(0.10)

1. Give a proof tree for SameGeneration(ewan, alice).
2. Compute the sets  $T_P^0, T_P^1, T_P^2, \dots$ . When is the fixed point reached?

**Exercise 8.2.** Consider databases that encode a labelled, directed graph by means of a ternary EDB predicate  $e$  (“edge”). The two parameters are the source and target nodes of the edge, while the third parameter is its label. For example, the edge  $n_1 \xrightarrow{a} n_2$  would be represented by the fact  $e(n_1, n_2, a)$ . Moreover, assume that only constants  $a$  and  $b$  are used as labels.

Can you express the following queries using Datalog?

1. “Which nodes in the graph are reachable from the node  $n$ ?”
2. “Are all nodes of the graph reachable from the node  $n$ ?”
3. “Does the graph have a directed cycle?”
4. “Does the graph have a path that is labelled by a palindrome?”  
(a palindrome is a word that reads the same forwards and backwards)
5. “Is the connected component that contains the node  $n$  2-colourable?”
6. “Is the graph 2-colourable?”
7. “Which pairs of nodes are connected by a path with an even number of  $a$  labels?”
8. “Which pairs of nodes are connected by a path with the same number of  $a$  and  $b$  labels?”
9. “Is there a pair of nodes that is connected by two distinct paths?”

**Exercise 8.3.** Consider a UCQ of the following form

$$(r_{11}(x) \wedge r_{12}(x)) \vee \dots \vee (r_{\ell 1}(x) \wedge r_{\ell 2}(x))$$

Find a Datalog query that expresses this UCQ. How many rules and how many additional IDB predicates does your solution use (depending on  $\ell$ )?

**Exercise 8.4.** Consider a Datalog query of the following form:

$$\begin{array}{lll} A_1(x) \leftarrow r_{11}(x) & \dots & A_\ell(x) \leftarrow r_{\ell 1}(x) \\ A_1(x) \leftarrow r_{12}(x) & \dots & A_\ell(x) \leftarrow r_{\ell 2}(x) \\ & & \mathbf{Ans}(x) \leftarrow A_1(x) \wedge \dots \wedge A_\ell(x) \end{array}$$

Find a UCQ that expresses this Datalog query. How many CQs does your solution contain (depending on  $\ell$ )?

**Exercise 8.5.** Show that  $T_P^\infty$  is the least fixed point of the  $T_P$  operator.

1. Show that it is a fixed point, i.e., that  $T_P(T_P^\infty) = T_P^\infty$ .
2. Show that every fixed point of  $T_P$  must contain every fact in  $T_P^\infty$ .