Humboldt-Universität zu Berlin
Lehrstuhl Logik in der Informatik
Prof. Dr. Nicole Schweikardt

14. Januar 2021

Einführung in die Datenbanktheorie

Wintersemester 2020/2021

Übungsblatt 8

Zu Bearbeiten bis zur Übungsstunde am *27. Januar 2021*

Aufgabe 1:

Finden Sie zwei Datalog-Programme P_1 und P_2 mit $\mathit{edb}(P_1) = \mathit{edb}(P_2)$ und $\mathit{idb}(P_1) = \mathit{idb}(P_2)$, so dass für $\mathbf{S} := \mathit{edb}(P_1) = \mathit{edb}(P_2)$ gilt:

(1) Es gibt eine Datenbank $\mathbf{J} \in \mathit{inst}(\mathbf{S})$ so dass $[\![P_1]\!](\mathbf{J}) \not\subseteq [\![P_2]\!](\mathbf{J})$, und

(2) es gibt ein $R \in \mathit{idb}(P_1)$, so dass für die Anfragen $Q_1 := (P_1, R)$ und $Q_2 := (P_2, R)$, sowie alle $\mathbf{I} \in \mathit{inst}(\mathbf{S})$ gilt:
$$[\![Q_1]\!](\mathbf{I}) \subseteq [\![Q_2]\!](\mathbf{I}).$$

Beachten Sie: Aussage (1) bedeutet, dass das *Uniforme Containment-Problem für Datalog-Programme* bei Eingabe von P_1 und P_2 die Ausgabe "nein" liefert; und Aussage (2) bedeutet, dass das *Query Containment Problem für Datalog-Anfragen* bei Eingabe von Q_1 und Q_2 die Ausgabe "ja" liefert.

Aufgabe 2:

In der Literatur wird die Semantik von Datalog oft durch den Fixpunkt der iterativen Anwendung des T_P-Operators definiert. Für jedes $\mathbf{J} \in \mathit{inst}(\mathit{sch}(P))$ ist dabei
$$T_P^0(\mathbf{J}) := \mathbf{J} \qquad \text{und} \qquad T_P^{i+1}(\mathbf{J}) := T_P(T_P^i(\mathbf{J})).$$

(a) Zeigen Sie, dass für jedes Datalog-Programm P, alle $i \in \mathbb{N}$ und jedes $\mathbf{I} \in \mathit{inst}(\mathit{edb}(P))$ gilt:
$$T_P^i(\hat{\mathbf{I}}) \;=\; S_P^i(\hat{\mathbf{I}}).$$

(b) Gilt sogar für alle $\mathbf{J} \in \mathit{inst}(\mathit{sch}(P))$, dass
$$T_P^i(\mathbf{J}) \;=\; S_P^i(\mathbf{J})\ ?$$

Beweisen Sie, dass Ihre Antwort korrekt ist.

Aufgabe 3:
Zeigen Sie das Lemma $\triangle$ aus der Vorlesung, d.h. zeigen Sie:

> Sei $\Sigma \subseteq \mathbf{dom}$. Sei $G = (V, \Sigma, S, P)$ eine kontextfreie Grammatik, für die gilt:
>
> *(i)* Es gibt keine Produktion der Form $X \to \epsilon$, für $X \in V$,
>
> *(ii)* Es gibt keine Produktion auf deren rechter Seite das Startsymbol S steht.
>
> Sei P_G das Datalog-Programm, welches für jede Produktion $A \to B_1 \cdots B_n$ aus G die Regel
>
> $$R_A(x_1, x_{n+1}) \leftarrow \tilde{B}_1, \ldots, \tilde{B}_n \quad \text{mit} \quad \tilde{B}_i := \begin{cases} E(x_i, b, x_{i+1}) & \text{falls } B_i = b \in \Sigma \\ R_X(x_i, x_{i+1}) & \text{falls } B_i = X \in V \end{cases}$$
>
> enthält. Sei $m \geq 1$ und seien $a_1, \ldots, a_m, b_1, \ldots, b_{m-1} \in \mathbf{dom}$. Dann gilt:
>
> $b_1 \cdots b_{m-1} \in L(G) \iff$ Es gibt einen Beweisbaum für das Faktum $R_S(a_1, a_m)$ bzgl P_G, dessen Blätter mit den Fakten $E(a_1, b_1, a_2), E(a_2, b_2, a_3), \ldots, E(a_{m-1}, b_{m-1}, a_m)$ markiert sind.

Aufgabe 4:
Zeigen Sie die im Beweis von Theorem 4.16 a) noch fehlende Rückrichtung, die Folgendes besagt:

$$Q_G \sqsubseteq Q_{G'} \quad \Longrightarrow \quad L(G) \subseteq L(G').$$