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

16. Dezember 2015

# Einführung in die Datenbanktheorie

Wintersemester 2015/2016

# Übungsblatt 8

**Bearbeitung:** *in den Übungen am 6./7. Januar 2016*

**Aufgabe 1:** **(12 + 13 Punkte)**

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:

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

**(b)** 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:

$$[\![Q_1]\!](\mathbf{I}) \subseteq [\![Q_2]\!](\mathbf{I}).$$

**Aufgabe 2:** **(25 Punkte)**

Zeigen Sie, dass für jede Dataloganfrage $Q := (P, R)$ gilt:

Die durch $Q$ definierte Anfragefunktion $[\![Q]\!]$ ist abgeschlossen unter adom($Q$)-Homomorphismen.

*Zur Erinnerung:*

Ein $C$-Homomorphismus (für $C \subseteq \mathbf{dom}$) ist eine Abbildung $h : \mathbf{dom} \to \mathbf{dom}$ mit $h_{|C} = \mathrm{id}$.

Eine Anfragefunktion $q$ ist abgeschlossen unter $C$-Homomorphismen, falls für alle $C$-Homomorphismen $h$ und alle Datenbanken $\mathbf{I}$ und $\mathbf{J}$ gilt:

$$\text{Falls } h(\mathbf{I}) \subseteq \mathbf{J}\text{ , so ist } h(q(\mathbf{I})) \subseteq q(\mathbf{J}).$$

**Aufgabe 3:** **(25 Punkte)**

Zeigen Sie das Lemma $\textcircled{\small\Delta}$ aus der Vorlesung, d.h. zeigen Sie:

Sei $\Sigma \subseteq \mathbf{dom}$. Sei $G = (\Sigma, V, S, P)$ 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

$$R_A(x_1, x_{n+1}) \leftarrow \tilde{B}_1, \ldots, \tilde{B}_n \quad \text{mit} \quad \tilde{B}_i := \begin{cases} E(x_i, a, x_{i+1}) & \text{falls } B_i = a \in \Sigma \\ R_X(x_i, x_{i+1}) & \text{falls } B_i = X \in V \end{cases}$$

enthält. Sei $m \geq 1$ und seien $a_1, \ldots, a_m, b_1, \ldots, b_{m-1} \in \mathbf{dom}$. Dann gilt:

$b_1 \cdots b_{m-1} \in L(G) \quad \Leftrightarrow \quad$ 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:** **(25 Punkte)**

Zeigen Sie, dass das folgende Auswertungsproblem für Boolsche Datalog-Anfragen (kombinierte Komplexität) EXPTIME-vollständig ist.

| Auswertungsproblem für Boolsche Datalog-Anfragen |
|---|
| *Eingabe:* Datalog-Anfrage $Q = (P, R)$, Datenbank $\mathbf{I}$. |
| *Frage:* Ist $[\![Q]\!](\mathbf{I}) \neq \emptyset$? |

Hierbei ist:

$$\text{EXPTIME} := \bigcup_{k \in \mathbb{N}} \text{DTIME}(2^{(n^k)}),$$

wobei $\text{DTIME}(2^{(n^k)})$ die Klasse aller Entscheidungsprobleme ist, die von einer deterministischen Turing-Maschine in Zeit $2^{(n^k)}$ gelöst werden können.

Hinweise zur Lösung der Aufgabe finden Sie auf Seite 387 in:

E. Dantsin, T. Eiter, G. Gottlob and A. Voronkov.
**Complexity and expressive power of logic programming.**
ACM Computing Surveys, Vol. 33, No. 3, pages 374-425. 2001.

***Frohe Weihnachten und einen Guten Rutsch ins neue Jahr!***