Thomas Haas Prof. Dr. Roland Meyer	**Theoretische Informatik 1** **Übungsblatt 4**	TU Braunschweig Wintersemester 2019/20

Ausgabe: 10.12.2019 Abgabe: 19.12.2019, 15:00

Geben Sie Ihre Lösungen bis Donnerstag, 19.12.2019, 15:00 Uhr, durch Einwerfen in die Übungskästen neben Büro IZ 343 ab. Geben Sie in Gruppen von 4 Personen ab.

Definition: Finite-state Transducer

Ein endlicher Transduktor (finite-state transducer) ist formal ein 6-Tupel $T = (Q, \Sigma, \Gamma, \rightarrow, q_0, Q_F)$ wobei

1. Q eine endliche Menge an Zuständen,
2. Σ und Γ endliche Alphabete,
3. $\rightarrow \subseteq Q \times (\Sigma \cup \{\tau\}) \times (\Gamma \cup \{\tau\}) \times Q$ eine Transitionsrelation,
4. q_0 und Q_F jeweils Startzustand und die Menge der Endzustände sind.

Im folgenden sind Notationen und wichtige Definitionen gegeben:

1. $(q, a, x, q') \in \rightarrow$ schreiben wir als $q \xrightarrow{a/x} q'$. Beim Lesen von a in q geht der Transduktor in q' über und produziert dabei x. Intuitiv, falls $a = \tau$, dann soll es sich um eine spontane Transition handeln, falls $x = \tau$, dann soll die Ausgabe leer sein.
2. $\rightarrow^* \subseteq Q \times (\Sigma \cup \{\tau\})^* \times (\Gamma \cup \{\tau\})^* \times Q$ bezeichne die reflexive, transitive Hülle von $\rightarrow$.
   - $q \xrightarrow{\varepsilon/\varepsilon}{}^* q$
   - $q \xrightarrow{w/o}{}^* q' \iff \exists q_0, q_1, \ldots, q_n : q = q_0 \xrightarrow{w_1/o_1} q_1 \xrightarrow{w_2/o_2} q_2 \xrightarrow{w_3/o_3} \ldots \xrightarrow{w_n/o_n} q_n = q_{|w|} = q'$.
3. $$\pi_\Sigma : (\Sigma \cup \{\tau\}) \rightarrow \Sigma^*, \pi_\Sigma(a) = \begin{cases} a, a \in \Sigma \\ \varepsilon, a = \tau \end{cases}$$
   ist eine Funktion, die einen Homomorphismus induziert, welcher die τ's in einem Wort löscht. Die τ's stehen für spontane Transitionen oder leere Ausgaben und sollen deshalb nicht *sichtbar* sein.
4. T induziert eine Relation $[T] \subset \Sigma^* \times \Gamma^*$ wie folgt:
   $$w[T]o \iff \exists w' \in (\Sigma \cup \{\tau\})^*, o' \in (\Gamma \cup \{\tau\})^* : q_0 \xrightarrow{w'/o'}{}^* q_f \in Q_F \quad \text{und} \quad \pi_\Sigma(w') = w, \pi_\Gamma(o') = o$$
   Wir sagen, dass $o \in \Gamma^*$ ein Output von $w \in \Sigma^*$ unter T ist.

5. T übersetzt nicht nur einzelne Wörter, sondern ganze Sprachen. Wir definieren für eine Sprache $\mathcal{L} \subseteq \Sigma^*$ die Übersetzung unter T als $T(\mathcal{L}) = \{o \in \Gamma^* \mid \exists w \in \mathcal{L} : w[T]o\} \subseteq \Gamma^*$.

Ein Transduktor kann als ein NFA mit spontanen Übergängen betrachtet werden, der nicht nur Eingabewörter akzeptiert, sondern dabei auch Wörter ausgibt; er übersetzt also Eingabewörter aus Σ^* zu Wörtern in Γ^*.

Transduktoren finden Anwendung in der Linguistik und der Verarbeitung von natürlichen Sprachen.

Aufgabe 1: Endliche Transduktoren [9 Punkte]

a) [3 Punkte] Konstruieren Sie einen Transduktor T, der in einem Wort $w \in \{a, b\}^*$ jedes zweite a löscht und jedes b verdoppelt. Geben Sie einen regulären Ausdruck für die Übersetzung von $(ab)^*$ unter T an.

   Ein Korrektheitsbeweis ist nicht nötig.

b) [3 Punkte] Konstruieren Sie einen Transduktor T, der in einem Wort $w \in \{a, b, c\}^*$ jedes Vorkommen von der Teilsequenz acb löscht und c's, die nicht in solch einer Sequenz auftauchen, beliebig vervielfachen (aber nicht löschen!) kann. Geben Sie einen regulären Ausdruck für die Übersetzung von $a^+(\varepsilon + c + cc)b^*$ unter T an.

   Ein Korrektheitsbeweis ist nicht nötig.

c) [3 Punkte] Wir nennen einen Transduktor deterministisch, wenn es in jedem Zustand zu jeder Eingabe **maximal eine** mögliche, und damit auch **eindeutige**, Transition gibt; diese darf auch spontan sein. Beispielsweise darf ein Zustand mit einer a-Transition keine weitere a-Transition oder spontane Transition besitzen. In beiden Fällen würde es 2 mögliche Transitionen geben, die der Transduktor beim Lesen eines a's nehmen könnte.

   Zeigen Sie, dass es nicht möglich ist Transduktoren zu determinisieren, d.h. es gibt nicht zu jedem Transduktor T einen deterministischen Transduktor T^{det} mit $T(\mathcal{L}) = T^{det}(\mathcal{L})$ für alle Sprachen $\mathcal{L} \in \Sigma^*$.

   Bemerkung: Obwohl bei der Erkennung regulärer Sprachen Nichtdeterminismus keine Rolle spielt (zu jedem NFA gibt es äquivalenten DFA), ist dies bei der Generierung von Sprachen mit Transduktoren nicht der Fall!

Aufgabe 2: Operationen durch Transduktoren [10 Punkte]

a) [4 Punkte] Sei $h : \Sigma^* \to \Gamma^*$ ein Homomorphismus zwischen Sprachen. Konstruieren Sie einen Transduktor T_h, sodass $T_h(\mathcal{L}) = h(\mathcal{L})$ für jede Sprache $\mathcal{L} \in \Sigma^*$ gilt. Beweisen Sie die Korrektheit ihrer Konstruktion.

b) [4 Punkte] Zeigen Sie nun, dass es auch einen Transduktor $T_{h^{-1}}$ gibt, sodass $T_{h^{-1}}(\mathcal{L}) = h^{-1}(\mathcal{L})$ für alle $\mathcal{L} \subseteq \Gamma^*$ gilt. Beweisen Sie die Korrektheit ihrer Konstruktion.

c) [2 Punkte] Zeigen Sie, dass es zu jeder regulären Sprachen M einen Transduktor T_M gibt mit $T_M(\mathcal{L}) = \mathcal{L} \cap M$.

Bemerkung: Hiermit haben Sie gezeigt, dass Transduktoren in der Lage sind, viele übliche Operationen auf Sprachen darzustellen. Ist eine Sprachklasse abgeschlossen unter Übersetzungen von Transduktoren, dann folgt nun direkt, dass diese auch unter den oben genannten Operationen abgeschlossen ist.

Aufgabe 3: Determinisierung ist teuer [7 Punkte]

In dieser Aufgabe wollen wir zeigen, dass sich manche Sprachen mit einem kleinen NFA beschreiben lassen, jeder DFA dafür jedoch zwangsweise riesig ist.

Für eine Zahl $k \in \mathbb{N}, k > 0$ sei

$$\mathcal{L}_{a@k} = \left\{ w \in \Sigma^* \;\middle|\; \text{Der } k\text{. letzte Buchstabe in } w \text{ ist } a \right\}$$

die Sprache der Wörter über $\Sigma = \{a, b\}$, deren k-ter Buchstabe von rechts ein a ist.

Beispielsweise ist $\mathcal{L}_{a@3} = \Sigma^*.a.(a \cup b).(a \cup b)$ die Sprache der Wörter, deren drittletzter Buchstabe a ist.

a) [2 Punkte] Zeigen Sie, wie man zu jedem $k \in \mathbb{N}, k > 0$ einen NFA $A_k = (Q, q_0, \rightarrow, Q_F)$ mit $\mathcal{L}(A_k) = \mathcal{L}_{a@k}$ konstruiert. Geben Sie diesen Automaten formal als Tupel an.

   Sie müssen die Sprachgleichheit nicht beweisen.

   Wie viele Zustände hat A_k?

b) [2 Punkte] Zeichnen Sie A_3 und bestimmen Sie seine Determinisierung A_3^{det} mittels der Rabin-Scott-Potenzmengenkonstruktion.

   Vergleichen Sie die Zustandsanzahl von A_3 und A_3^{det}.

c) [3 Punkte] Sei $k \in \mathbb{N}, k > 0$ beliebig. Beweisen Sie, dass es für $\mathcal{L}_{a@k}$ keinen DFA B mit weniger als 2^k Zuständen gibt, so dass $\mathcal{L}(B) = \mathcal{L}_{a@k}$ gilt.

   Hinweis: Gehen Sie wie folgt vor:

   1. Angenommen es gäbe $B = (Q', q'_0, \rightarrow', Q'_F)$ mit $\mathcal{L}(B) = \mathcal{L}_{a@k}$ und $|Q'| < 2^k$.
   2. Betrachten Sie die Menge Σ^k der Wörter der Länge k. Wie viele solcher Wörter gibt es?
   3. Betrachten Sie zu jedem Wort $w \in \Sigma^k$ den (eindeutigen) Zustand q_w, in dem DFA B ist, nachdem er w gelesen hat.
   4. Leiten Sie nun einen Widerspruch her.

Aufgabe 4: Äquivalenzrelationen [7 Punkte]

Es sei $\equiv \subseteq \Sigma^* \times \Sigma^*$ eine Äquivalenzrelation auf Wörtern. Wie üblich schreiben wir $u \equiv v$ (statt $(u, v) \in \equiv$), um auszudrücken, dass u und v gemäß $\equiv$ äquivalent sind.

a) [2 Punkte] Beweisen Sie formal die folgenden grundlegenden Eigenschaften von Äquivalenzrelationen:

- Jedes Wort ist in seiner Äquivalenzklasse enthalten: $u \in [u]_\equiv$.
- Die Äquivalenzklassen von äquivalenten Wörtern sind gleich: $u \equiv v \implies [u]_\equiv = [v]_\equiv$.
- Die Äquivalenzklassen von nicht-äquivalenten Wörtern sind disjunkt: $u \not\equiv v \implies [u]_\equiv \cap [v]_\equiv = \varnothing$.

b) [2 Punkte] Es sei $\mathcal{L} \subseteq \Sigma^*$ und $\equiv_\mathcal{L}$ die aus der Vorlesung bekannte Nerode-Rechtskongruenz mit

$$u \equiv_\mathcal{L} v \quad \text{gdw.} \quad \forall w \in \Sigma^*\colon u.w \in \mathcal{L} \iff v.w \in \mathcal{L}\,.$$

Beweisen Sie, dass $\equiv_\mathcal{L}$ tatsächlich eine Äquivalenzrelation und Rechtskongruenz ist. Letzteres bedeutet, dass für alle u, v mit $u \equiv_\mathcal{L} v$ und alle $x \in \Sigma^*$ gilt: $u.x \equiv_\mathcal{L} v.x$.

c) [2 Punkte] Es sei $A = (Q, q_0, \to, Q_F)$ ein DFA. Die Relation $\equiv_A \subseteq \Sigma^* \times \Sigma^*$ ist definiert durch:

$$u \equiv_A v \quad \text{gdw.} \quad \exists q \in Q\colon q_0 \xrightarrow{u} q \text{ und } q_0 \xrightarrow{v} q.$$

Zeigen Sie, dass $\equiv_A$ eine Äquivalenzrelation ist.

d) [1 Punkt] Ist $\equiv_A$ aus Aufgabenteil c) auch eine Äquivalenzrelation, wenn A ein NFA ist? Begründen Sie Ihre Antwort!

Aufgabe 5: (Bonus) Mächtigkeit von Transduktoren [6 Bonuspunkte]

In Aufgabe 2 haben Sie gezeigt, dass eine unter Übersetzungen von Transduktoren abgeschlossene Sprachklasse auch unter homomorphen Bildern, Urbildern und dem Schnitt mit regulären Sprachen abgeschlossen ist.

Zeigen Sie nun, dass dies auch in umgekehrter Richtung gilt. Das heißt, eine Sprachklasse, die unter den oben aufgeführten Operationen abgeschlossen ist, ist auch unter Übersetzungen von Transduktoren abgeschlossen.

Bemerkung: Sie müssen also zeigen, dass Sie die Übersetzung einer beliebigen Sprache $\mathcal{L}$ unter einem Transduktor T mit Hilfe von den drei oben genannten Operation darstellen können.