Hinweise für die Tutorien NACH der 4. Vorlesung (11.11.09)

1 Formale Sprachen

- aus der Vorlesung:
  - formale Sprache: $L \subseteq A^*$
  - Produkt: $L_1 \cdot L_2 = \{w_1 w_2 \mid w_1 \in L_1 \land w_2 \in L_2\}$
  - Potenzen: $L^0 = \{\varepsilon\}$ und $L^{i+1} = L^i \cdot L$
  - Konkatenationsabschluss:

$$L^+ = \bigcup_{i=1}^{\infty} L^i \qquad \text{und} \qquad L^* = \bigcup_{i=0}^{\infty} L^i$$

- Beispiele machen:
  - formale Sprache L_I der legalen Zahlen vom Typ **int**:
    - Versuch: $A = \{\mathtt{0}, \ldots, \mathtt{9}\}$
      $L_I = A^+$. (nicht A^*)
    - Was fehlt? Z. B. das Minuszeichen;
      besser: $\{\varepsilon, \mathtt{-}\}A^+$
    - Was ist mit Präfix `0x`? ich weiß es nicht
  - formale Sprache L_V der legalen Variablennamen in Java:
    - Versuch: $A = \{\mathtt{\_}, \mathtt{a}, \ldots, \mathtt{z}, \mathtt{A}, \ldots, \mathtt{Z}\}$, $B = A \cup \{\mathtt{0}, \ldots, \mathtt{9}\}$
      $L_V = A \cdot B^*$.
    - es fehlen die Umlaute, ...
    - Was ist noch falsch? z. B. Schlüsselwörter (**if**, ...) sind als Variablennamen verboten.
      also eher sowas wie $L_V = (A \cdot B^*) \setminus \{\mathtt{if}, \mathtt{class}, \ldots\}$ da könnte man jetzt alle endlich vielen Schlüsselwörter aufzählen
      aber wenn nur endlich viele Wörter verboten sind, geht es im Prinzip ohne Mengendifferenz
  - formale Sprache L aller Wörter über $A = \{\mathtt{a}, \mathtt{b}\}$, in denen nirgends das Teilwort `ab` vorkommt.
    - Das kann man auch positiv formulieren: In den erlaubten Wörtern müssen, wenn überhaupt, erst alle `b` kommen und danach, wenn überhaupt alle `a`
    - also $L = \{\mathtt{b}\}^*\{\mathtt{a}\}^*$
  - Bitte darauf achten, dass nicht Wörter und Sprachen durcheinander geworfen werden:
    - `abb` ist etwas anderes als $\{\mathtt{abb}\}$.
    - Und $\{\mathtt{abb}\}^*$ gibt es, aber $\mathtt{abb}^*$ **gibt es nicht** (bis jetzt).
  - $L_1 = \{\mathtt{a}^n \mid n \in \mathbb{N}_0\}$ und $L_2 = \{\mathtt{b}^n \mid n \in \mathbb{N}_0\}$
    Achtung: $L_1 L_2 = \{\mathtt{a}^k \mathtt{b}^m \mid k \in \mathbb{N}_0 \land m \in \mathbb{N}_0\}$ die **Exponenten können verschieden sein!**

- allgemeines zu Mengen:
  - generell: wissen die Studenten, was Mengendifferenz ist?
  - Darauf hinweisen: $\{1, 2, 3\} \cup \{2, 3, 4\} = \{1, 2, 3, 4\}$
    kein Element kann „mehrfach vorkommen“.
    Wer so was wie $\{1, 2, 3, 2, 3, 4\}$ schreibt, steht im dringenden Verdacht, etwas noch nicht verstanden zu haben.

- Man beweise: $L^* \cdot L = L^+$
  - Wie beweist man, dass zwei Mengen gleich sind?
  - Zum Beispiel, indem man zeigt, dass $\subseteq$ und $\supseteq$ gelten.
  - Also:
    - $\subseteq$:
      Wenn $w \in L^* \cdot L$, dann $w = w'w''$ mit $w' \in L^*$ und $w'' \in L$.
      Also existiert ein $i \in \mathbb{N}_0$ mit $w' \in L^i$.
      Also $w = w'w'' \in L^i \cdot L = L^{i+1}$.
      Da $i + 1 \in \mathbb{N}_+$, ist $L^{i+1} \subseteq L^+$, also $w \in L^+$.

* $\supseteq$: Wenn $w \in L^+$, dann existiert ein $i \in \mathbb{N}_+$ mit $w \in L^i$.
  Da $i \in \mathbb{N}_+$ ist $i = j+1$ für ein $j \in \mathbb{N}_0$,
  also ist für ein $j \in \mathbb{N}_0$: $w \in L^{j+1} = L^j \cdot L$.
  also $w = w'w''$ mit $w' \in L^j$ und $w'' \in L$.
  Wegen $L^j \subseteq L^*$ ist $w = w'w'' \in L^* \cdot L$.

2 Kontextfreie Grammatiken

- aus der Vorlesung: für alle Alphabete A und alle $x \in A$ Funktionen $N_x : A^* \to \mathbb{N}_0$, die wie folgt festgelegt sind:

$$N_x(\varepsilon) = 0$$
$$\forall y \in A : \forall w \in A^* : N_x(yw) = \begin{cases} 1 + N_x(w) & \text{falls } y = x \\ N_x(w) & \text{falls } y \neq x \end{cases}$$

  $N_x(w)$ gibt an, wie oft x in w vorkommt. Fragen, ob das klar ist.

- aus der Vorlesung: $G = (N, T, S, P)$ mit
  - $N \cap T = \emptyset$
  - $S \in N$ und
  - $P \subseteq N \times V^*$, P endlich, wobei $V = N \cup T$ sei.
  - Produktionen schreiben wir meist in der Form $X \to w$
  - mehrere mit gleicher linker Seite zusammengefasst: $X \to w_1 \mid w_2 \mid \cdots \mid w_k$

- Man arbeite mit $G = (\{X\}, \{\mathtt{a}, \mathtt{b}\}, X, \{X \to \varepsilon \mid \mathtt{a}X \mid \mathtt{b}X\})$
  - Was kann man alles ableiten? ε, $\mathtt{a}$, $\mathtt{b}$, $\mathtt{aa}$, ...
  - aha: alle Wörter überhaupt: $L(G) = \{\mathtt{a}, \mathtt{b}\}^*$

- Gibt es auch eine Grammatik G mit $L(G) = \emptyset$?
  - suchen lassen ...
  - z. B. $(\{X\}, \{\mathtt{a}, \mathtt{b}\}, X, \{X \to X\})$.
  - wir haben sogar leere Produktionenmenge zugelassen: $(\{X\}, \{\mathtt{a}, \mathtt{b}\}, X, \{\})$ tuts auch.
  - allerdings: leere Alphabete haben wir verboten, also $(\{X\}, \{\}, X, P)$ geht *nicht*.

- Man arbeite mit $G = (\{X\}, \{\mathtt{(}, \mathtt{)}\}, X, \{X \to XX \mid \mathtt{(}X\mathtt{)} \mid \varepsilon\}$
  - man mache Beispielableitungen
    * erste einfache wie $X \Rightarrow \mathtt{(}X\mathtt{)} \Rightarrow \mathtt{((}X\mathtt{))} \Rightarrow \mathtt{(((}X\mathtt{)))} \Rightarrow \mathtt{((((}X\mathtt{))))} \Rightarrow \mathtt{(((())))}$ oder
    * $X \Rightarrow XX \Rightarrow XXX \Rightarrow XXXX \Rightarrow XXXXX$ und dann irgendwie weiter
  - Welche Wörter w sind ableitbar?
    * anschaulich: ableitbar sind genau die „*wohlgeformten Klammerausdrücke*“
    * jedenfalls gleich viele $\mathtt{(}$ und $\mathtt{)}$: $N_{\mathtt{(}}(w) = N_{\mathtt{)}}(w)$
    * Das ist aber nur notwendig aber nicht hinreichend für Ableitbarkeit (man diskutiere diese Adjektive), denn $\mathtt{)(}$ ist z. B. nicht ableitbar.
    * zusätzliche Eigenschaften? erst mal raten/ nachdenken/ rumprobieren lassen
    * aha: für jedes Präfix (es heißt *das* Präfix) v eines $w \in L(G)$ gilt: $N_{\mathtt{(}}(v) \geq N_{\mathtt{)}}(v)$
      Das kann man sich gerade noch klar machen; aber der Beweis, dass man damit eine notwendige und hinreichende Bedingung für Ableitbarkeit hat, also eine Charakterisierung der Klammerausdrücke, ist wohl zu schwierig; ich sehe jedenfalls auf Anhieb keine vernünftige Erklärung.

- Man arbeite mit $G = (\{X\}, \{\mathtt{(}, \mathtt{)}\}, X, \{X \to \mathtt{(}X\mathtt{)}X \mid \varepsilon\})$.
  - siehe da: auch damit sind genau die wohlgeformten Klammerausdrücke ableitbar

  - Man mache sich klar, warum …

- Und dann auch Grammatiken konstruieren *lassen*, z. B. für die folgenden formalen Sprachen über dem Alphabet $T = \{\mathtt{a}, \mathtt{b}\}$.

  - die Menge aller Wörter über T, in denen irgendwo das Teilwort `baa` vorkommt,
    z. B. so: $(\{X, Y\}, T, X, P)$ mit $P = \{X \to Y\mathtt{baa}Y, Y \to \mathtt{a}Y|\mathtt{b}Y|\varepsilon\}$
  - die Menge aller Wörter $w \in T^*$ mit der Eigenschaft, dass für alle Präfixe v von w gilt: $|N_\mathtt{a}(v) - N_\mathtt{b}(v)| \leq 1$.
    * Man überlege sich erst mal, welche Struktur Wörter der Länge 2, 4, … haben: wenn ich das richtig sehe: $\{\mathtt{ab}, \mathtt{ba}\}^*$
    * Also leistet die Grammatik $(\{X, Y\}, T, X, P)$ mit $P = \{X \to \mathtt{ab}X|\mathtt{ba}X|\mathtt{a}|\mathtt{b}|\varepsilon\}$ das Gewünschte.

- **Achtung**: bitte nicht aus Versehen mit Grammatiken bzw. formalen Sprachen vom Aufgabenblatt 5 rumspielen

3 Reflexiv-transitive Hülle

- Standard-Definitionen aus der Vorlesung

  - für $R \subseteq M_1 \times M_2$ und $S \subseteq M_2 \times M_3$:
    $S \circ R = \{(x, z) \in M_1 \times M_3 \mid \exists y \in M_2 : (x, y) \in R \wedge (y, z) \in S\}$
  - $\mathrm{Id}_M = \{(x, x) \mid x \in M\}$
  - $R^0 = \mathrm{Id}_M$ und $\forall i \in \mathbb{N}_0 : R^{i+1} = R \circ R^i$
  - $R^* = \bigcup_{i=0}^{\infty} R^i$

- z. B. in der Vorlesung offen gelassen:

  - Es sei R eine beliebige Relation und S eine Relation, die reflexiv und transitiv ist. Wenn $R \subseteq S$, dann ist sogar $R^* \subseteq S$.
  - Man beweise das, indem man durch vollständige Induktion zeigt: Für alle $i \in \mathbb{N}_0$: Wenn $R \subseteq S$, dann $R^i \subseteq S$.
  - Wenn man eine Relation hin malt: Elemente $x, y \in M$ als Punkte und einen Pfeil von x nach y, falls xRy (Infixnotation wird in der Vorlesung eingeführt): Wie sieht das Bild aus, wenn die Relation reflexiv ist? Schlingen. Wie, wenn sie transitiv ist? (schwieriger zu beschreiben; nur Beispiele ansehen; Wenn man man einen Zyklus dabei hat: jeder mit jedem verbunden)