3 Formale Sprachen

3.1 Mengen, Teil 2

Da formale Sprachen Mengen sind, sicherheitshalber nochmal ein bisschen was zum Mengenbegriff:

- 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.
- $M \cup \{\} = M$
- $M \cap \{\} = \{\}$
- Wissen die Studenten, was Mengendifferenz ist?
  - Beispiel: $\{1,2,3\} \setminus \{2,3,4\} = \{1\}$
  - $M_1 \setminus M_2 = \{x \mid x \in M_1 \land x \notin M_2\}$

3.2 Formale Sprachen, Beispiele

- Def: $L \subseteq A^*$
- Bitte darauf achten, dass nicht Wörter und Sprachen durcheinander geworfen werden:
  - `abb` ist etwas anderes als $\{\texttt{abb}\}$.
  - Und $\{\texttt{abb}\}^*$ gibt es, aber $\texttt{abb}^*$ **gibt es nicht** (bis jetzt).
- formale Sprache aller Schlüselwörter in Java: eine endliche Sprache
  $L = \{\texttt{class}, \texttt{if}, \texttt{int}, \dots\}$
- formale Sprache L aller Wörter über $A = \{\texttt{a}, \texttt{b}\}$, in denen nirgends das Teilwort `ab` vorkommt.
  - das kann man z. B. so hinschreiben: $L = \{\texttt{a}, \texttt{b}\} \setminus \{w_1\texttt{ab}w_2 \mid w_1, w_2 \in \{\texttt{a}, \texttt{b}\}^*\}$
  - man überlege, was dann noch übrig bleibt
  - positiv formuliert: In den erlaubten Wörtern muss erst ein beliebiges Wort (evtl. leer) nur aus `b` kommen und danach ein beliebiges Wort (evtl. leer) nur aus `a`.
  - man kann also auch schreiben $L = \{w_1w_2 \mid w_1 \in \{\texttt{b}\}^* \land w_2 \in \{\texttt{a}\}^*$

3.3 Produkt oder Konkatenation formaler Sprachen

3.3.1 Produkt von Sprachen

- Def: $L_1 \cdot L_2 = \{w_1 w_2 \mid w_1 \in L_1 \wedge w_2 \in L_2\}$
- Beispiel: noch mal: formale Sprache L aller Wörter über $A = \{\mathtt{a}, \mathtt{b}\}$, in denen nirgends das Teilwort `ab` vorkommt.
  kann man jetzt so schreiben: $L = \{\mathtt{b}\}^*\{\mathtt{a}\}^*$
- Beispiel: Menge aller Wörter über A außer dem leeren: $A \cdot A^*$, denn jedes nichtleere Wort hat ein erstes Symbol und dahinter kommt ein beliebiges (evtl auch leeres) Wort
- formale Sprache L_I der legalen Zahlen vom Typ **int**:
  - Versuch: $A = \{\mathtt{0}, \ldots, \mathtt{9}\}$
    $L_I = A \cdot A^*$
  - Was fehlt? jedenfalls das Minuszeichen;
    besser: $\{\varepsilon, \mathtt{-}\} \cdot A \cdot A^*$
- formale Sprache L_V der legalen Variablennamen in Java:
  - Versuch: $A = \{\_, \mathtt{a}, \ldots, \mathtt{z}, \mathtt{A}, \ldots, \mathtt{Z}\}$, $B = A \cup \{\mathtt{0}, \ldots, \mathtt{9}\}$
    $L_V = A \cdot B^*$.
  - es fehlen die Umlaute, ...
  - Was ist noch falsch? z. B. Schlüsselwörter (**if**, ...) sind als Variablennamen verboten.
    also eher sowas wie $L_V = (A \cdot B^*) \smallsetminus \{\mathtt{if}, \mathtt{class}, \ldots\}$
    Mitteilung: da könnte man jetzt alle endlich vielen Schlüsselwörter aufzählen, aber wenn nur endlich viele Wörter verboten sind, geht es im Prinzip ohne Mengendifferenz
- $L_1 = \{\mathtt{a}^n \mid n \in \mathbb{N}_0\}$ und $L_2 = \{\mathtt{b}^n \mid n \in \mathbb{N}_0\}$
  Achtung: $L_1 L_2 = \{\mathtt{a}^k \mathtt{b}^m \mid k \in \mathbb{N}_0 \wedge m \in \mathbb{N}_0\}$ die **Exponenten können verschieden sein!**
  hier steht nichts anderes als $\{\mathtt{a}\}^*\{\mathtt{b}\}^*$

3.3.2 Potenzen von L

- Def: $L^0 = \{\varepsilon\}$ und $L^{i+1} = L^i \cdot L$
- Beispiel: $L = \{\mathtt{a}\}^*\{\mathtt{b}\}^*$
  dann enthält man z. B.

- $L^0 = \{\varepsilon\}$
- $L^1 = L = \{\varepsilon, \texttt{aabbbbb}, \texttt{aaaaab}, \texttt{aaaa}, \texttt{bbbbbbbb}, \dots\}$
- $L^2 = \{\varepsilon, \texttt{aabbbbaaaaab}, \texttt{aaaaabab}, \texttt{aaaaa}, \texttt{bbbbbbb}, \dots\}$
- usw.

3.4 Konkatenationsabschluss einer formalen Sprache

3.4.1 Kokatenationsabschluss

- Def:

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

- Beispiel: $L = \{\texttt{a}\}^*\{\texttt{b}\}^*$
  - Man mache sich klar: $L^* = \{\texttt{a}, \texttt{b}\}^*$, also alles
  - das geht z. B. so: (die Studenten möglichst selber drauf kommen lassen) zerhacke beliebiges aber festes $w \in \{\texttt{a}, \texttt{b}\}^*$ an allen Stellen, wo Teilwort `ba` vorkommt, zwischen dem `b` und dem `a`. Die entstehenden Teilwörter sind aus L.
- 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 $w = w'w'' \in L^* \cdot L$.

3.5 alte Klausuraufgabe

- Zum Schluss (und zur Verdeutlichung von Konkatenation und Schnitt zweier Sprachen) ist es vielleicht ganz nett, eine alte Klausuraufgabe rechnen zu lassen. Ich schlage einfach mal Aufgabe 1 aus der September 2009 Klausur vor (kann aber auch gerne eine andere genommen werden):

  http://gbi.ira.uka.de/archiv/2008/k-sep09.pdf

- Ich weiss leider nicht, ob die Studenten schon mit “modulo n“ vertraut sind. An dieser Stelle vielleicht nachfragen und bei Bedarf ein paar erklärende Sätze zur Restdivision sprechen.

- Man sollte außerdem die Aufgabenstellung etwas abändern und nicht nach regulären Ausdrücken fragen (das Kennen die Studenten noch nicht), sondern nach der beschriebenen Sprache in der etwas unschönen Mengenschreibweise.