Prof. Dr. Hans-Jörg Kreowski **Mai 2007**

Formale Sprachen: DNA Computing

Sticker-Systeme

Ein *Sticker-System* $\gamma = (V, \rho, A, D)$ besteht aus einem Alphabet V, einer symmetrischen Relation $\rho \subseteq V \times V$, einer endlichen Menge von *Axiomen* $A \subseteq LR_\rho(V)$ und einer endlichen Menge von *Bausteinen* $D \subseteq W_\rho(V) \times W_\rho(V)$.

Ein *Berechnungsschritt* ist gegeben durch $x \Longrightarrow u \circ x \circ v$ für $(u, v) \in D$ und $x \in LR_\rho(V)$.

Eine *Berechnung* ist eine Folge von Berechnungsschritten, die mit einem Axiom beginnt:

$$x_0 \Longrightarrow x_1 \Longrightarrow \ldots \Longrightarrow x_k \text{ mit } x_0 \in A.$$

Eine *Berechnung* heißt *vollständig*, wenn $x_k \in WK_\rho(V)$.

Die *Molekülsprache* von γ enthält alle vollständigen Moleküle, die durch Berechnungen entstehen:

$$LM(\gamma) = \{w \in WK_\rho(V) \mid x \overset{*}{\Longrightarrow} w, x \in A\}.$$

Die *erzeugte Sprache* von γ enthält alle oberen Stränge von Molekülen der Molekülsprache:

$$L(\gamma) = \{w \in V^* \mid \begin{bmatrix} w \\ w' \end{bmatrix} \in LM(\gamma), w' \in V^*\}.$$

Da Berechnungsschritte durch Anbauen von Bausteinen am linken und rechten Ende von Molekülen mit Überhängen definiert sind, können die Moleküle während einer Berechnung nur monoton wachsen. Das ist signifikant, weil daraus folgt, dass Sticker-Sprachen kontextsensitiv sind. Das lässt sich folgendermaßen einsehen:

Sei $w \in V^*$ ein Wort, das auf das Band einer Turing-Maschine geschrieben wird. Dann kann diese Turing-Maschine daneben eine Berechnung simulieren, denn dazu muss sie ein Axiom von endlich vielen schreiben und dann wiederholt links und rechts Bausteine von endlich vielen anbauen. Das berechnete Molekül wird dabei solange aufgebaut, bis die Länge von w erreicht ist. Dann wird Vollständigkeit geprüft, und im positiven Fall wird jeweils das untere Symbol eines Symbolpaares gelöscht, so dass nur der obere Strang übrigbleibt. Schließich vergleicht die Turing-Maschine das Ergebnis mit w. Diese Turing-Maschine erkennt die Sprache $L(\gamma)$ und braucht dafür doppelt so viel Band wie das zu erkennende Eingabewort lang ist. Damit handelt es sich um einen sogenannten linear beschränkten Automaten (LBA). Von dieser Automatenklasse ist bekannt, dass sie genau die kontextsensitiven Sprachen erkennt, die mit den Sprachen übereinstimmen, die von monotonen Chomsky-Grammatiken erzeugt werden.

Bezeichne ASL die Klasse aller allgemeinen Sticker-Sprachen und CS die Klasse aller kontextsensitiven Sprachen, dann lässt sich das erzielte Ergebnis knapp und einprägsam notieren:

$$ASL \subseteq CS.$$

Beachte, dass die Sticker-Sprache $L(\gamma_1)$ nicht kontextfrei ist, so dass sich die Einbettung von ASL in die Chomsky-Hierarchie nicht verbessern lässt.

Es folgen drei weitere Beispiele, die die Funktionsweise und Leistungsfähigkeit von Sticker-Systemen illustrieren.

1. $\gamma_2 = (\{a,b\}, \{(a,a),(b,b)\}, \{\left[\begin{smallmatrix} ab \\ ab \end{smallmatrix}\right]\}, \{(\lambda, \left(\begin{smallmatrix} ab \\ \lambda \end{smallmatrix}\right)), (\lambda, \left(\begin{smallmatrix} \lambda \\ ab \end{smallmatrix}\right))\})$ erzeugt
   $L(\gamma_2) = \{(ab)^n \mid n \geq 1\}$, eine typisch reguläre Sprache.

2. $\gamma_3 = (\{a,b\}, \{(a,a),(b,b)\}, \{\left[\begin{smallmatrix} ab \\ ab \end{smallmatrix}\right]\}, \{(\left(\begin{smallmatrix} a \\ \lambda \end{smallmatrix}\right), \left(\begin{smallmatrix} b \\ \lambda \end{smallmatrix}\right)), (\left(\begin{smallmatrix} \lambda \\ a \end{smallmatrix}\right), \left(\begin{smallmatrix} \lambda \\ b \end{smallmatrix}\right))\})$ erzeugt
   $L(\gamma_3) = \{a^n b^n \mid n \leq 1\}$, eine typisch lineare Sprache, die nicht regulär ist.

3. $\gamma_4 = (\{a,b\}, \{(a,a),(b,b)\}, A_4, D_4)$ mit

   $$A_4 = \{\begin{bmatrix} a \\ a \end{bmatrix}, \begin{bmatrix} aa \\ aa \end{bmatrix}, \begin{bmatrix} aaa \\ aaa \end{bmatrix}, \begin{bmatrix} aba \\ aba \end{bmatrix}, \begin{bmatrix} aaaa \\ aaaa \end{bmatrix}, \begin{bmatrix} abaa \\ abaa \end{bmatrix}, \begin{bmatrix} a \\ a \end{bmatrix}^{aaa}, \begin{bmatrix} a \\ a \end{bmatrix}^{baa}, \begin{bmatrix} aba \\ aba \end{bmatrix}^{b}\}$$

   $$D_4 = \{(\lambda, {}_x\begin{bmatrix} a \\ a \end{bmatrix}^{aaa}), (\lambda, {}_x\begin{bmatrix} a \\ a \end{bmatrix}^{baa}), (\lambda, {}_x\begin{bmatrix} aba \\ aba \end{bmatrix}^{b}) \mid x \in \{a,b\}\} \cup$$

   $$\{\ (\lambda, {}_{xy}\begin{bmatrix} ba \\ ba \end{bmatrix}^{ba}) \mid x,y \in \{a,b\}\} \cup$$
   $$\{\ (\lambda, {}_{xyz}\begin{bmatrix} a \\ a \end{bmatrix}^{aaa}) \mid x,y,z \in \{a,b\}\} \cup$$
   $$\{\ (\lambda, {}_x\begin{bmatrix} a \\ a \end{bmatrix}), (\lambda, {}_x\begin{bmatrix} aa \\ aa \end{bmatrix}), (\lambda, {}_x\begin{bmatrix} aaa \\ aaa \end{bmatrix}), (\lambda, {}_x\begin{bmatrix} aba \\ aba \end{bmatrix}) \mid x \in \{a,b\}\} \cup$$

   $$\{\ (\lambda, {}_{xy}\begin{bmatrix} \lambda \\ \lambda \end{bmatrix}), (\lambda, {}_{xy}\begin{bmatrix} ba \\ ba \end{bmatrix}) \mid x,y \in \{a,b\}\} \cup$$

   $$\{\ (\lambda, {}_{xyz}\begin{bmatrix} \lambda \\ \lambda \end{bmatrix}), (\lambda, {}_{xyz}\begin{bmatrix} a \\ a \end{bmatrix}), (\lambda, {}_{xyz}\begin{bmatrix} aa \\ aa \end{bmatrix}), (\lambda, {}_{xyz}\begin{bmatrix} aaa \\ aaa \end{bmatrix}), \mid x,y,z \in \{a,b\}\}.$$