FGI 1

Automaten, Formale Sprachen, Berechenbarkeit

Christopher Habel & Matthias Jantzen

Kap 25

(Überblick zu
Automaten und Formale Sprachen)

Matthias Jantzen

Was sind Homomorphismen?

Definition 82:

Eine Abbildung h von einer Halbgruppe $(H, \odot)$ auf eine andere Halbgruppe $(G, \circledast)$ wird genau dann Homomorphismus genannt, wenn $\forall x, y \in H : h(x \odot y) = h(x) \circledast h(y)$ gilt. Ein Homomorphismus ist also eine strukturerhaltende Abbildung, die außer den Elementen auch noch die Operationen auf einander abbildet.

Sei $\log : I\!R \to I\!R$ die bekannte Logarithmusfunktion, dann gilt:

$$\log(x \cdot y) = \log(x) + \log(y)$$

Unterhalbgruppe und Untermonoid

Definition 84:

Für eine Teilmenge $M \subseteq \Sigma^*$ zum Alphabet Σ waren M^+ und M^* wie folgt erkärt:

$$M^+ := \bigcup_{i \geq 1} M^i \text{ mit } M^1 := M \text{ und } M^{i+1} := M^i \cdot M$$

$$M^* := M^+ \cup \{\epsilon\}$$

- M^+ heißt **von M erzeugte Unterhalbgruppe**.
- M^*heißt **von M erzeugtes Untermonoid**.

freies Erzeugendensystem

Definition 85:

- M heißt **Erzeugendensystem** von M^+ bzw. von M^*. Ist das Erzeugendensystem M endlich, so heißen M^+ bzw. M^* **endlich erzeugt**.
- M^+ bzw. M^* heißt **von M frei erzeugt** genau dann, wenn für $\forall i, k, r \in I\!N$ $\forall u_i, v_i \in M \setminus \{\epsilon\}$ aus

  $$u_1 \cdot u_2 \cdot \ldots \cdot u_k = v_1 \cdot v_2 \cdot \ldots \cdot v_r$$

  stets $k = r$ und $u_i = v_i$ folgt.

Substitution

Wir hatten bisher gesehen, dass die Familie $\mathcal{R}eg$ der regulären Mengen abgeschlossen ist gegenüber den Operatoren: Vereinigung ($\vee$), (Komplex-)Produkt, Sternbildung, Komplement, Mengendifferenz und Durchschnitt ($\wedge$). Das Wechseln des Alphabets durch Austausch einzelner Symbole ist offensichtlich auch keine Schwierigkeit. Man ersetze die entsprechenden Symbole in dem die Menge beschreibenden rationalen Ausdruck. Diese Idee wird nun verallgemeinert, indem für einzelne Symbole nun ganze Sprachen *substituiert* werden:

Substitution bei Wörtern

Definition 87:

Eine **Substitution** ist ein Homomorphismus $s : \Sigma^* \longrightarrow 2^{\Gamma^*}$, wobei Σ und Γ Alphabete sind und s folgende Eigenschaften besitzt:

1. Für jedes $a \in \Sigma$ ist $s(a) \subseteq \Gamma^*$ definiert.
2. $s(\epsilon) := \{\epsilon\}$.
3. $\forall u, v \in \Sigma^* : s(uv) = s(u) \cdot s(v)$

Ist $s(a)$ regulär (bzw. endlich) für jedes $a \in \Sigma$, dann heißt s eine **reguläre** bzw. eine **endliche** Substitution.

Substitution bei Sprachen

noch Definition 87:

Für eine Sprache $L \subseteq \Sigma^*$ werde s kanonisch (in natürlicher Weise) erweitert durch:

$$s(L) := \bigcup_{w \in L} s(w)$$

Sei $\Sigma := \{a, b\}$, $\Gamma := \{0, 1\}$ sowie $s : \Sigma^* \to \Gamma^*$ definiert durch: $s(a) := \{1\}^*$, $s(b) := \{00\}$. Dann gilt:

$$s(aabb) = \{1\}^*\{1\}^*\{0000\} = \{1\}^*\{0000\}$$

$$s(\{a^n b^n \mid n \in I\!N\}) = \{1\}^*\{(00)^n \mid n \in I\!N\} = \{1\}^*\{00\}^*$$

reguläre Mengen sind Substitutionsabgeschlossen

Satz 53:

Die Familie der regulären Mengen ist gegenüber regulären Substitutionen abgeschlossen.

Beweis:

Jede reguläre Menge R wird durch einen rationalen Ausdruck dargestellt. Ersetzt man nun jedes Symbol a in diesem Ausdruck durch den rationalen Ausdruck der $s(a)$ beschreibt, so ergibt sich ein rationaler Ausdruck für $s(R)$.

die Familie Cf ist Substitutionsabgeschlossen

Satz 54:

Die Familie der kontextfreien Sprachen ist gegenüber kontextfreien Substitutionen abgeschlossen.

Beweis:

Sei G eine *CFG* für $L := L(G)$ und G_a disjunkte Grammatiken für jedes Terminal a mit $s(a) = L(G_a)$.
Ersetzt man nun jedes Symbol a in rechten Seiten der Produktionen von G durch das Startsymbol der zugehörigen *CFG*, so entsteht eine neue *CFG* für $s(L)$.

inverser Homomorphismus

Definition 88:

Seien Σ und Γ endliche Alphabete sowie $h : \Sigma^* \longrightarrow \Gamma^*$ ein Homomorphismus. Die zu h inverse Relation, genannt **inverser Homomorphismus**, kann auch als totale Funktion $h^{-1} : \Gamma^* \longrightarrow 2^{\Sigma^*}$ angesehen werden. Hierbei definiert h^{-1} für jedes $w \in \Gamma^*$ die Menge aller möglichen Urbilder von w durch:

$$h^{-1}(w) := \{v \in \Sigma^* \mid h(v) = w\}.$$

Diese Menge kann daher auch leer sein. Auf formale Sprachen erweitert ergibt sich kanonisch:

$$h^{-1}(L) := \bigcup_{w \in L} h^{-1}(w)$$

Beispiel 1

Sei $h : \{a,b,c\}^* \longrightarrow \{x,y\}^*$ definiert durch

$$\begin{array}{rcl} a & \mapsto & xyx, \\ b & \mapsto & xy, \\ c & \mapsto & yx. \end{array}$$

Dann ist

$$\begin{array}{rcl} h^{-1}(xy) & = & \{b\}, \\ h^{-1}(xx) & = & \emptyset \quad und \\ h^{-1}(xyxyx) & = & \{ac, ba\}. \end{array}$$

Beispiel 2

Sei $f : \{a,b,c\}^* \longrightarrow \{x,y\}^*$ definiert durch

$$\begin{array}{rcl} a & \mapsto & x, \\ b & \mapsto & y, \\ c & \mapsto & \epsilon. \end{array}$$

Dann ist

$$\begin{array}{rcl} f^{-1}(\{\epsilon\}) & = & \{c\}^*, \\ f^{-1}(\{y\}) & = & \{c\}^*\{b\}\{c\}^* \quad und \\ f^{-1}(\{x,y\}^*) & = & \{a,b,c\}^*. \end{array}$$

Reg und Cf und inverse Homorphismen:

Satz 55:

Jede der Familien $\mathcal{Cf}$ und $\mathcal{Reg}$ ist gegenüber Anwendung von inversen Homomorphismen abgeschlossen.

Auf die Beweise verzichten wir hier!

Überblick über Inhalte von FGI-1 zu $AFSB$

Funktionen vs. Relationen

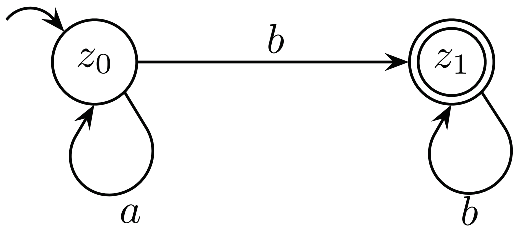
- Funktionen sind **eindeutig**,
- Relationen brauchen nicht eindeutig zu sein!

Wo ist der Unterschied wichtig?

- In der Automatentheorie:
  - Determinismus (*DFA*, *DPDA*),
  - Nichtdeterminismus (*NFA*, *NPDA*)
- *allgemeiner*: In der Graphentheorie;
- überall, wo keine Eindeutigkeit gefordert ist (z.B. $\leq, \subseteq$).

Was ist ein endlicher Automat?

- *DFA* $A = (Q, \Sigma, \delta, z_0, F)$
  - δ meist grafisch dargestellt durch Zustandsübergangs-diagramm, z.B.:



  - oder als Übergangstabelle mit $F := \{z_1\}$

δ	a	b
z_0	z_0	z_1
z_1		z_1

Eigenschaften

- *DFA* vs. *NFA*
  - Zu jedem *NFA* existiert äquivalenter *DFA*.
  - **Idee:** Merke in den *Zuständen des Potenzautomaten*, wo man sich im *NFA* aufhalten könnte.
  - Endzustände sind diejenigen, in denen im *NFA* ein Endzustand erreichbar war.
- Vollständigkeit
- weitere Eigenschaften: (fast) buchstabierend

Abschlusseigenschaften (Typ-3)

Für reguläre Mengen R_1 und R_2, auch regulär:

- Vereinigung $R_1 \cup R_2$
- Komplexprodukt $R_1 \cdot R_2$
- Sternbildung R_1^*
- Durchschnitt $R_1 \cap R_2$
- Komplement $\Sigma^* \setminus R_1$
- Reversal R_1^{rev}
- (inverse) Homomorphismen $h(R_1),\ h^{-1}(R_1)$
- reguläre Substitution $\sigma(R_1)$

reguläre Ausdrücke

Sei Σ ein Alphabet.

Reguläre „Grund-“Ausdrücke sind:

- $\emptyset$ ist regulärer Ausdruck $\qquad M_\emptyset = \emptyset = \{\}$
- a ist regulärer Ausdruck für alle $a \in \Sigma$ $\qquad M_a = \{a\}$

Sind A und B reguläre Ausdrücke, dann sind auch folgende Ausdrücke regulär:

- $(A + B)$ $\qquad M_{(A+B)} = (M_A \cup M_B)$
- $A \cdot B$ $\qquad M_{A \cdot B} = M_A \cdot M_B$
- $(A)^*$ $\qquad M_{(A)^*} = (M_A)^*$

reguläre Ausdrücke \ vs. reguläre Mengen

Reguläre Ausdrücke beschreiben genau die regulären Mengen!

- Durch endliche Automaten beschriebene Sprachen lassen sich auch durch reguläre Ausdrücke beschreiben.

  [$R^k_{i,j}$-Konstruktion von Kleene]

- Durch reguläre Ausdrücke beschriebene Mengen sind auch durch endliche Automaten darstellbar.

  [Abschlusseigenschaften]

kontextfreie Sprachen

... lassen sich beschreiben durch

- Kontextfreie Grammatiken $G = (\Sigma, N, P, S)$

  $A \longrightarrow w$ $N \times (N \cup \Sigma)^*$

- Chomsky-Normalform-Grammatiken $V_N \times (V_N^2 \cup V_T)$

  $A \longrightarrow BC$ oder $A \longrightarrow a$

- Greibach-Normalform-Grammatiken $N \times \Sigma \cdot N^*$

  $A \longrightarrow aA_1A_2 \cdots$

- Kellerautomaten $A = (Z, \Sigma, \Gamma, K, Z_{start}, Z_{end})$

ein *DPDA* muss immer ein Symbol auf dem Keller ansehen! Spezialfall: *RLG*s äquivalent zu *NFA*s

Ableitung(-sbaum)

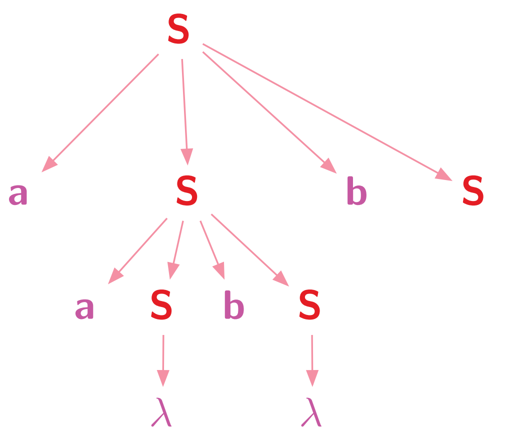
Grammatik mit Regeln: $S \longrightarrow aSbS \mid \lambda$

Ableitungen können geschrieben werden:

- linear:

  $S \Rightarrow aSbS \Rightarrow aaSbSbS \Rightarrow aabSbS \Rightarrow aabbS \Rightarrow aabb$

- als Baum:



Ableitungsbäume + Chomsky-Normalform

$\Rightarrow$ Pumping-Lemma

Teilklassen der kontextfreien Sprachen

Echte Teilmengen der kontextfreien Sprachen sind:

- Durch eindeutige kontextfreie Grammatiken erzeugte Sprachen: $w \in L(G)$ hat **genau einen Ableitungsbaum** bzw. **genau eine Linksableitung**.

- Durch deterministische Kellerautomaten akzeptierte Sprachen.

  Beispiel:

  $L := \{wcw^{rev} \mid w \in \{a, b\}^*\}$ ist deterministisch kontextfrei und eindeutig:

  $S \longrightarrow aSa \mid bSb \mid c$ ist eindeutige Grammatik für L.

- Die regulären Mengen. [Konstruktion einer CFG aus einem FA]

Akzeptierungsbedingungen

Ein Wort w wird von einem *endlichen Automaten* A akzeptiert, gdw. es eine Erfolgsrechnung von A auf w gibt. $(z_0, w) \stackrel{*}{\longmapsto} (z_e, \epsilon)$ mit $z_0 \in Q_{start}$ und $z_e \in F$

Bei Kellerautomaten: zwei unterschiedliche Bedingungen.

- mit **Endzustand** akzeptierte Sprache $L(A)$ $(q_0, w, \perp) \stackrel{*}{\longmapsto} (p, \epsilon, v)$ mit $q_0 \in Q_{start}$ und $p \in F$
- mit **leerem Keller** akzeptierte Sprache $N(A)$ $(q_0, w, \perp) \stackrel{*}{\longmapsto} (q, \epsilon, \epsilon)$ mit $q_0 \in Q_{start}$ und $q \in Q$

Beide Akzeptierbedingungen sind für *PDA*'s äquivalent!

DBDA's akzeptieren mit leerem Keller nur **präfixfreie** Sprachen!

Abschlusseigenschaften (Typ-2)

	kontextfrei	det. kontextfrei
Vereinigung	•	
Durchschnitt		
Durchschnitt mit regulären Mengen	•	•
Komplexprodukt	•	
Sternbildung	•	
Komplement		•
kontextfreie Substitution	•	
Homomorphismen	•	
inverse Homomorphismen	•	

Entscheidbarkeiten (Typ 2)

	kontextfrei	det. kontextfrei
Leerheit	•	•
Universalität		•
Endlichkeit	•	•
Wortproblem	•	•
Inklusion		
Äquivalenz		•
Eindeutigkeit		•

Nicht entscheidbar sind dagegen (unter anderem):

- **Ein-/Mehrdeutigkeit kontextfreier Grammatiken**
- $L_1 \cap L_2 = \emptyset$ **für kontextfreie Sprachen** L_1 **und** L_2

formale Schreibweisen

Es ist wichtig, z.B. Mengen formal korrekt zu notieren:

„Die Menge L aller Wörter, die zwei d's enthalten.“

- Über welchem Alphabet? $\Sigma := \{c, d\}$
- **Mindestens** zwei oder **genau** zwei d's?
- Darstellungsmöglichkeiten:
  - $L = L(A)$ für
    $A = (\{0,1,2\}, \{c,d\}, \{(0,c,0), (0,d,0), (0,d,1),$
    $(1,c,1), (1,d,1), (1,d,2), (2,c,2), (2,d,2)\}, \{0\}, \{2\})$
  - $L = L(G)$ für $G = (\{c,d\}, \{S,A\}, P, S$
    mit $P := \{(S, AdAdA), (A, cA), (A, dA), (A, \epsilon)\}$
  - $L = M_R$ für $R = (c+d)^* d (c+d)^* d (c+d)^*$

noch wenige Tage bis zur Klausur…

Klausurtermin:

27. Juli 2009

09:00 - 12:00

Audimax 1

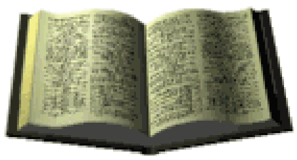
Einlass von 09:00 bis 09:20 Uhr.
Jede bzw. jeder meldet sich zuerst vorne am Pult,
dort wird auf die Plätze „verteilt“!

und nun…

(ver)zweifeln Sie nicht
an sich selbst…

entspannen Sie
sich…

Lesen Sie nur wenig
zum Thema…



Viel Erfolg!