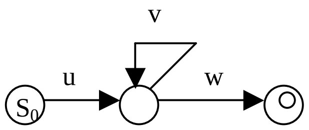
Pumping Lemma

1.) Automat mit Zyklus



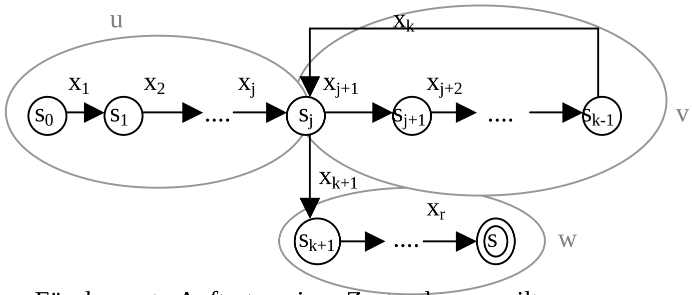
2.) reguläre Ausdrücke

*-Operator ist die einzige Möglichkeit, unendliche Sprachen zu beschreiben
----→ .... L* ....----→ sehr lange Wörter haben viele Wiederholungen des selben Teilworts

Beweis:

Sei R eine reguläre Sprache. Zu R gibt es einen endlichen deterministischen Akzeptor $A=(X,S,\delta,s_0,F)$ mit $L(A)=R$. Wähle $n:=|S|$. Sei $z \in R$ mit $|z| \geq n$. Dann gibt es mindestens einen Zustand, der bei Eingabe von z zweimal durchlaufen wird:

$z = x_1..x_r, \quad r \geq n, \quad x_i \in X$ für $i = 1,...,r$.



Für das erste Auftreten eines Zustands $s_j=s_k$ gilt:
$u:= x_1\text{-}x_j$, $v:= x_{j+1}...x_k \neq \varepsilon$ $\quad |uv| = |x_1...x_k| \leq n$ wegen „erstem Auftreten"
$w:= x_{k+1}....x_r$

Offenbar ist $\delta(s_0,u) = \delta(s_0,uv) = \delta(s_0,uv^i)$, $i \geq 0$, also $\delta(s_0,uv^iw) = \delta(s_0,uvw) = s \in F \Rightarrow uv^iw \in R$, $i \geq 0$

Beachte: Pumping Lemma ist <u>keine</u> Äquivalenzaussage, nur eine Implikation:
L regulär $\Rightarrow$ L erfüllt PL
aber nicht $\Leftarrow$

Typische Anwendung:

Nutze PL um zu zeigen, daß eine Sprache L nicht regulär ist. Aber auch das geht oft nicht, denn es gibt auch Sprachen, die erfüllen das PL, sind aber dennoch nicht regulär:

$L=\{c^m a^n b^n \mid m,n \geq 0\} \cup \{a,b\}^*$
v ← kann gepumpt werden
Pumpvarianten: für uvw → $v=c^i$ → pumpe v und lande in L
→ $v=a^jb^j$ → pumpe v und lande in L

Bsp.: 1.) $L=\{a^n b^n \mid n \geq 0\}$ ist nicht regulär

2.) $L=\{0^p \mid p$ ist Primzahl$\} \subset \{0\}^*$

Annahme: L ist regulär. Dann erfüllt L das PL. Wähle n wie in Satz W. Sei r eine Primzahl r > n. Sei $z = 0^r \in L$, dann gibt es eine Zerlegung z = uvw mit $|uv| \leq n$, $u=0^t$, $v=0^i$, $t \neq 0$. Mit 0^r ist auch $0^{r+it} \in L$ für alle $i \geq 0$. Folglich sind alle Zahlen r+it Primzahlen. Nach spätestens t Zahlen kommt also stets eine Primzahl. Setze i = r, dann ist r+rt eine Primzahl $\Rightarrow$ r(1+t) ist eine Primzahl, andererseits sind r und 1+t Faktoren von r(1+t) → Widerspruch
$\Rightarrow$ L ist nicht regulär.

3.) $L=\{0^m \mid m$ ist Quadratzahl$\}$ ist nicht regulär

Annahme: L regulär. Dann gibt es ein $n \in N$, so daß sich jedes Wort der Form 0^m, $m \geq n$, m Quadratzahl, zerlegen läßt, in z=uvw mit entsprechenden Eigenschaften:
$v \neq \varepsilon$, $|uv| \leq n$, $uv^iw \in L$, $i \geq 0$.

Wähle speziell $z = 0^{n^2}$ und betrachte zugehörige Zerlegung z = uvw. Dann ist offenbar wegen der

Bedingung des PL: $1 \leq |v| \leq |uv| \leq n$.

Ferner ist für i=2: $uv^2w \in L$

Andererseits: $n^2 = |z| = |uvw| < |uv^2w| \leq n^2+n < n^2+2n+1 = (n+1)^2$

uvvw außerhalb der Sprache

n^2 n

$|uv^2w| < (n+1)^2$

$\notin L$ Widerspruch

4.7. Reguläre Grammatiken

Bisher:
- Prozesse des Erkennens von → Automaten
- Charakterisierungen → PL, Abschlußeigenschaften, algebraisch
- Beschreibungen → reg. Ausdruck
- Erzeugungsprozesse → Grammatiken

<Satz> → <Subjekt> <Prädikat> <Objekt>
<Subjekt> → Katze
<Prädikat> → jagt
<Objekt> → Maus

Startsymbol : <Satz>
Nichtterminalsymbol : <..>
Terminalsymbol : Hund, Katze,...

Ableitungsprozeß:
<Satz> → <Subjekt> <Prädikat> <Objekt> → Katze <Prädikat> <Objekt> → Katze jagt <Objekt> → Katze jagt Maus

Def. X : 1.) Eine Grammatik G beschreibt man durch ein 4-Tupel $G=(N,T,P,\sigma)$ wobei gilt:
- N nichtleere, endl. Menge von Nichtterminalsymbolen
- T nichtleere, endl. Menge von Terminalsymbolen
- $N \cap T = \varnothing$
- $P \subseteq \{(\alpha,\beta) \mid \alpha \in (N \cup T)^*, \beta \in (N \cup T)^*\}$ endl. Menge der Produktion, Produktionsregeln, Ableitungsregeln
  [Bemerkung: Statt (α, β) schreiben wir meist $\alpha \rightarrow_G \beta$ oder $\alpha \rightarrow \beta$]
- $\sigma \in N$ Startsymbol

2.) Seien $v,w \in (N \cup T)^*$. v ist ableitbar aus w (in Zeichen $w \rightarrow_G^* v$), wenn es Wörter gibt $v_1,..,v_k$, $u_1,..,u_k$, $z_1,..,z_k$, $\alpha_1,..,\alpha_{k-1}$, $\beta_1,..,\beta_{k-1} \in (N \cup T)^*$ gibt, so daß gilt:

$v_1=w$, $v_k=v$, $v_i=u_i\alpha_i z_i$, $v_{i+1}=u_i\beta_i z_i$ für i=1,..,k-1 und $(\alpha_i, \beta_i) \in P$ für i=1,..,k-1

Bsp.: $v_1 = u_1\alpha_1 z_1 \rightarrow u_1\beta_1 z_1 = u_2\alpha_2 z_2 \rightarrow u_2\beta_2 z_2=...$

3.) Die von G erzeugte Sprache L(G) ist definiert durch $L(G)=[w \in T^* \mid \alpha \rightarrow_G^* w\}$

Beispiele für Def. X:

1.) $L_1=\{a^n b^n \mid n \in N_0\}$ $G_1=(N_1,T_1,P_1,\sigma_1)$ mit $T_1=\{a,b\}$, $N_1=\{\sigma_1\}$, $P_1=\{\sigma_1 \rightarrow a\sigma_1 b, \sigma_1 \rightarrow \varepsilon\}$
$\sigma_1 \rightarrow a\sigma_1 b \rightarrow aa\sigma_1 bb \rightarrow ... \rightarrow a^n\sigma_1 b^n \rightarrow a^n b^n$

2.) korrekte Klammerung von begin und end
$G_2=(N_2,T_2,P_2,\sigma_2)$ mit $T_2=\{begin,end\}$, $N_2=\{\sigma_2\}$, $P_2=\{\sigma_2 \rightarrow begin\ \sigma_2\ end\ \sigma_2 , \sigma_2 \rightarrow \varepsilon\}$

```
begin
    begin
    end
    begin
    end
end
```

$\sigma_2 \rightarrow$ begin σ_2 end $\sigma_2 \rightarrow$ begin σ_2 end $\rightarrow$ begin begin σ_2 end σ2 end $\rightarrow$
begin begin end σ_2 end $\rightarrow$ begin begin end begin σ_2 end σ_2 end $\rightarrow$
begin begin end begin end end

Typisierung von Grammatiken und der Bezug zu speziellen Sprachklassen erfolgt über Einschränkung der Form der Produktionen.

Def. Y: Eine Grammatik $G=(N,T,P,\sigma)$ heißt rechtslineare Grammatik, wenn gilt :
Für alle $(\alpha,\beta) \in P$ gilt $\alpha \in N$ und $\beta \in T^*$ oder $\beta = \beta' B$ mit $\beta' \in T^*$, $B \in N$
[analog Linkslinearität : $\beta = \beta' B$ ersetzen durch $\beta = B\beta'$]

Bsp.: $L= \{a\}^* \cdot \{b\}^*$, $G=(N,T,P,\sigma)$ mit $N=\{\sigma,\sigma'\}$, $T=\{a,b\}$,
$P=\{\sigma \rightarrow a\sigma, \sigma \rightarrow \sigma', \sigma' \rightarrow b\sigma' , \sigma' \rightarrow \varepsilon\}$,
L(G)=L

Satz Z : Sei X ein Alphabet. Zu jeder regulären Sprache $L \subseteq X^*$ gibt es eine rechtslineare Grammatik G mit L=(G) und umgekehrt.

Beweis: „⇒“

$L \subseteq X^*$ regulär. Dann gibt es einen endlichen Akzeptor $A=(X,S,\delta,s_0,F)$ mit L=L(A).
Idee: Konstruiere Grammatik aus der Zustandsübergangsfuktion („δ wird zur Grammatik“)
Sei $G=(N,T,P,\sigma)$ definiert durch $N:=S$, $T:=X$, $\sigma:=s_0$,
$P:=\{s \rightarrow xs' \mid \forall s,s' \in S, x \in X: \delta(s,x)=s'\} \cup \{s \rightarrow \varepsilon \mid \forall s \in F\}$

zu zeigen : L(G)=L
Sei $w \in L(G)$, $w=x_1..x_k$, $x_i \in X$ $\Leftrightarrow_{Def.}$ $s_0 \rightarrow_G^* w$ $\Leftrightarrow$ $s_0 \rightarrow x_1 s_2 \rightarrow x_1 x_2 s_2 \rightarrow ... \rightarrow$
$x_1..x_k s_k \rightarrow x_1..x_k$ $\Leftrightarrow$ $\forall i \in \{0,..,k-1\}$: $\delta(s_i,x_{i+1})=s_{i+1}$ und $s_k \in F$ $\Leftrightarrow$ $\delta^*(s_0,w)=s_k \in F$
$\Leftrightarrow$ $w \in L(A)=L$