Erstellt von:
(Matrikelnummer: 702899)

Index:

4.6 Das *Pumping-Lemma* für reguläre Sprachen:

Das *Pumping-Lemma* ist ein zentrales Hilfsmittel um zu entscheiden bzw. um nachzuweisen, dass eine Sprache <u>nicht</u> regulär ist.

Satz W :

Es sei X ein Alphabet, dann existiert zu jeder regulären Sprache $R \subseteq X^*$ ein $n \in IN$, so dass für alle Wörter $z \in R$ mit $|z| \geq u$ eine Zerlegung $z = uvw$ mit $u, v, w \in X^*, v \neq \varepsilon$ und $|uv| \leq n$ existiert, für die gilt: $uv^i w \in R$ für alle $i \geq 0$.

Formal:

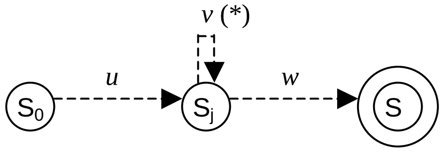
$$\forall R \in \mathrm{Re}g \;\; \exists n \in IN \;\; \forall z \in R, |z| \geq n \;\; \exists uvw \in X^* : z = uvw \wedge v \neq \varepsilon \wedge |uv| \leq n : \; \forall n \in IN_0 : uv^i w \in R \; .$$

Zugrundeliegende Idee des *Pumping-Lemma* :

Es existiert eine Pumpstelle in allen regulären Wörter z ab einer gewissen Länge n.

① *Endliche Akzeptoren* : Um unbeschränkte Wörter zu erkennen, muss ein Akzeptor in Zyklen laufen, wobei man diese Zyklen mehrfach durchlaufen kann.

Graphische Veranschaulichung:



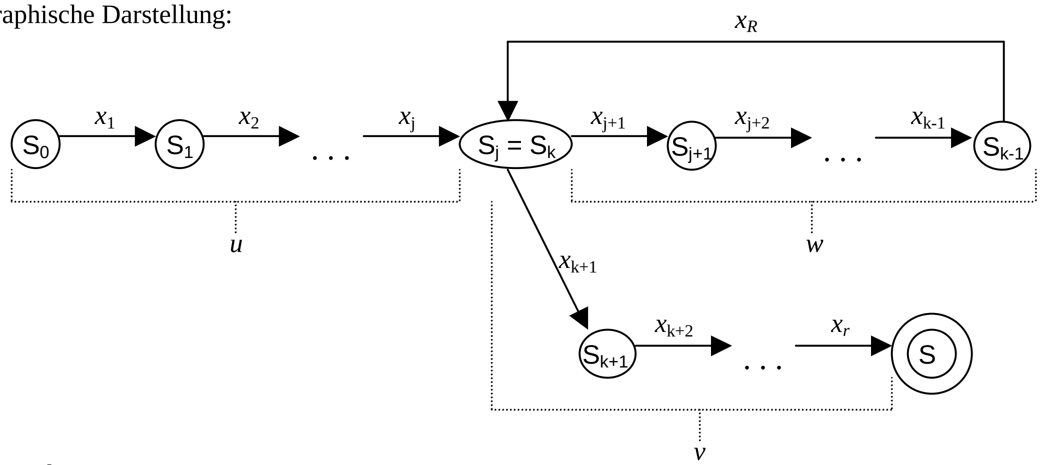
② *Reguläre Ausdrücke* : Ein * -Operator ermöglicht einige Methoden eine unendliche Sprache zu beschreiben. Dies ergibt, dass sehr lange Wörter viele Wiederholungen desselben Teilwortes enthalten. $w_1 \dashrightarrow L^* \dashrightarrow w_n \qquad n \in IN$

Beweis des *Pumping-Lemma* :

Es sei R eine reguläre Sprache und zu R existiert ein endlicher deterministischer Akzeptor $A = (X, S, \delta, s_0)$ mit $L(A) = R$.
Nun wähle man $n := |s|$, wobei $z \in R$ und $|z| \geq n$ ist.
Dann gibt es mindestens einen Zustand, welcher bei der Eingabe von z zweimal durchlaufen wird, also :

$$z = x_1,...,x_r \text{ , wobei gilt: } r \geq n \text{ und } x_i \in X \text{ für } i = 1,...,r \text{ .}$$

Graphische Darstellung:



Nun erkennt man:
Für das <u>erste</u> Auftreten eines Zustandes $s_j = s_k$ gilt:

$$u := x_1 \ldots x_j \text{ und } v := x_{j+1} \ldots x_R \neq \varepsilon \text{ .}$$

Damit gilt für das <u>erste</u> Auftreten : $|uv| = |x_1 \ldots x| \leq n$. Damit ist $w := x_{k+1} \ldots x_r$.

Daraus lässt sich nun folgern:
Offenbar ist $\delta(s_0, n) = \delta(s_0, uv) = \delta(s_0, uv^i)$ mit $i \geq 0$, woraus folgt $\delta(s_0, uv^i w) = \delta(s_0, uvw) = s \in F$
$\Rightarrow uv^i w \in R$ mit $i \geq 0$.

□

Hinweise zum *Pumping-Lemma* :

① Das *Pumping-Lemma* ist keine Äquivalenzaussage. Es ist nur eine Implikation:

Ist eine Sprache L regulär, dann erfüllt L das *Pumping-Lemma*.

Der umgekehrte Weg ist nicht möglich.

② Eine Typische Anwendung des *Pumping-Lemma* ist um zu zeigen, dass eine Sprache L nicht regulär ist. Aber das *Pumping-Lemma* kann oft nicht Angewendet werden, da es Sprachen gibt, welche die Bedingungen des *Pumping-Lemma* erfüllen, aber dennoch nicht regulär sind.

Beispiel :

Man nehme an es existiert eine nicht reguläre Sprache L mit $L = \{c^m a^n b^n \mid m, n \geq 0\} \cup \{a, b\}^*$,
dann existieren folgende *Pump*möglichkeiten für L :

L unterteilt in uvw: 1) $v = c^i$, mit $i \geq 0 \Rightarrow v$ wird gepumpt und das Wort bleibt Element von L.
2) $v = a^i b^i$, mit $i \geq 0 \Rightarrow a, b$ wird gepumpt und das Wort bleibt Element von L.

Beispiele zum *Pumping-Lemma* :

1) Man definiere eine nicht reguläre Sprache $L = \{O^P \mid p \text{ ist eine Primezahl}\}$.
Annahme : L ist regulär. $\Rightarrow$ Dann muss L das *Pumping-Lemma* erfüllen.
Daraus folgt diese Beweisführung:

Man wähle ein n nach dem *Satz W* und es sei r eine Primezahl mit $r > n$.
Des weiteren sei $z = O^r \in L$.
$\Rightarrow$ Dann existiert eine Zerlegung $z = uvw$ mit $|uv| \leq n$, $u = O^s$, $v = O^t$ mit $s, t \neq 0$.
Für O^r gilt auch $O^{r+i\cdot t} \in L$ für alle $i \geq 0$.
$\Rightarrow$ Folglich sind alle Zahlen $r + i \cdot t$ Primezahlen. Nach spätestens t Zahlen kommt also immer eine Primezahl. Nun setze man $i = r$, dann ist $r + r \cdot t$ eine Primezahl.

$\Rightarrow$ $r \cdot (1+t)$ ist eine Primezahl, andererseits sind r und $1 + t$ Faktoren von $r \cdot (1+t)$. ↯
$\Rightarrow$ L ist nicht regulär.

2) Es sei die Sprache $L = \{O^m \mid m \text{ ist Quadratzahl}\}$ nicht regulär.
Man gehe auch hier von der Annahme aus, dass L regulär ist.
Die Beweisführung lautet dann folgendermaßen:

Unter der Bedingung, dass L regulär ist, muss es ein $n \in IN$ existieren, so dass sich jedes Wort zu der Form O^m mit $m \geq n$ und m Quadratzahl , sich in die Form $z = uvw$ zerlegen lässt, mit den entsprechenden Eigenschaften:

$v \neq \varepsilon$, $|uv| \leq n$, $uv^i w \in L$ mit $i \geq 0$.

Nun wähle man speziell: $z = 0^{n^2}$
und betrachte zugleich die Zerlegung $z = uvw$.
Daraus folgt wegen der Bedingung des *Pumping-Lemma* : $1 \leq |v| \leq |uv| \leq n$.

Ferner ist für $i = 2$: $uv^2 w \in L$, anderseits soll gelten: $n^2 = |z| = |uvw| < |uv^2 w| \leq n^2 + n < n^2 + n + 1 = (n+1)^2$

$\Rightarrow \underbrace{(nv^2 w)}_{\notin L} < (n+1)^2$ ↯ $\Rightarrow L$ ist nicht regulär !

4.7 Reguläre Grammatiken

Allgemeine Grammatiken :

Zusammenfassung des bisher behandelten Stoffes :

- Prozesse des Erlernen von Sprachen → Automaten
- Charakterisierungen → Pumping-Lemma, Abschlusseigenschaften, algebraisch
- Beschreibungen → reguläre Ausdrücke
- Erzeugungprozesse → *Grammatiken*

Beispiel einer allgemeinen Grammatik :

Produktions- / Ableitungsregeln :
Es handelt sich hierbei um jene Regeln, nach denen die verschiedenen Symbole verwendet bzw. ersetzt werden dürfen.

<Satz> → <Subjekt> <Prädikat> <Objekt>

<Subjekt> → Hund <Objekt> → Katze
<Subjekt> → Katze <Objekt> → Maus

<Prädikat> → beißt
<Prädikat> → jagt

Erläuterung der verschiedenen Symbolen:

<...>	: Nichtterminalsymbole
Hund, Katze, beißt, ...	: Terminalsymbole
<Satz>	: Startsymbol

Ableitungsprozess :
Beispiel einer Anwendung der Produktionsregeln dieser Grammatik.

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

Definition X :

① Eine *Grammatik G* beschreibt man durch ein Quadrupel $G = (N, T, P, \sigma)$ wobei folgendes gilt:
- N ist die nichtleere endliche Menge von *Nichtterminalsymbolen*.
- T ist die nichtleere endliche Menge von *Terminalsymbolen*.
- $N \cap T = \emptyset$, Es gibt keine *Terminalsymbolen* die gleichzeitig *Nichtterminalsymbolen* sind und umgekehrt.
- $\sigma \in N$ ist das Startsymbol.
- $P \subseteq \{(\alpha,\beta) \mid \alpha,\beta \in (N \cup T)^*\}$ ist die endliche Menge der *Produktionen* oder auch *Produktions*- oder *Ableitungsregeln*.

Bemerkung :
Statt (α,β) schreibt man auch $\alpha \xrightarrow[G]{} \beta$ oder auch $\alpha \longrightarrow \beta$, wenn der Bezug zur Grammatik eindeutig ist.

② Es seien $v, w \in (N \cup T)^*$. Dann sei des weiteren v <u>ableitbar</u> aus w
(Formal dargestellt: $w \xrightarrow[G]{}^* v$, bzw. vereinfacht $w \longrightarrow^* v$ oder $w \longrightarrow v$) ,

wenn Wörter der Art $v_1,...,v_k$; $u_1,...,u_k$; $z_1,...,z_k$; $\alpha_1,...,\alpha_{k-1}$; $\beta_1,...,\beta_{k-1} \in (N \cup T)^*$
existieren, dann 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 alle $i = 1,...,k-1$.

Darstellung: $v_1 = u_1\alpha_1 z_1 \to u_1\beta_1 z_1 = u_2\alpha_2 z_2 \to u_2\beta_2 z_2 = u_3\alpha_3 z_3 \to ... \to v$.

③ Die von einer *Grammatik G* <u>erzeugte</u> <u>Sprache</u> $L(G)$ ist definiert durch:
$$L(G) = \{w \in T^* \mid a \xrightarrow[G]{}^* w\}.$$

Beispiele für allgemeine Grammatiken nach der Definition X :

1) Man nehme die Sprache $L_1 = \{a^n b^n \mid n \in IN\}$, für welche die Grammatik $G_1 = (N_1, T_1, P_1, \sigma_1)$ mit $N_1 = \{\sigma_1\}$, $T_1 = \{a, b\}$ und $P_1 = \{\sigma_1 \to a\sigma_1 b, \sigma_1 \to \varepsilon\}$.
In diesem Fall sieht der Ableitungsprozess folgendermaßen aus:
$\sigma_1 \to a\sigma_1 b \to aa\sigma_1 bb \to ... \to a^n \sigma_1 b^n \to a^n b^n$, woraus folgt: $L(G_1) = L_1$.

2) Man definiere für die korrekte Klammerung durch *begin* und *end* in einem Programm folgende Grammatik $G_2 = (N_2, T_2, P_2, \sigma_2)$ mit $N_2 = \{\sigma_2\}$, $T_2 = \{begin, end\}$ und $P_2 = \{\sigma_2 \to begin\,\sigma_2 end\,\sigma_2, \sigma_2 \to \varepsilon\}$.
Der Ableitungsprozess besitzt folgendes Aussehen:
$\sigma_2 \to begin\,\sigma_2 end\,\sigma_2 \to begin\,\sigma_2 end \to begin\,begin\,\sigma_2 end\,\sigma_2 end \to begin\,begin\,end\,\sigma_2 end$
$\to begin\,begin\,end\,begin\,\sigma_2 end\,\sigma_2 end \to begin\,begin\,end\,begin\,\sigma_2 end\,end \to begin\,begin\,end\,begin\,end\,end$

```
     begin
       begin
⇒      end
       begin
       end
     end
```

Reguläre Grammatiken :

Die Typisierung von Grammatiken und der Bezug zu speziellen Sprachklassen erfolgt über die Einschränkung der Form der Produktion.

Definition Y :

Eine Grammatik $G = (N, T, P, \sigma)$ heißt *rechtslineare Grammatik*, wenn gilt

Für alle $(\alpha, \beta) \in P$ existiert ein $\alpha \in N$ und ein $\beta \in T^*$ oder ein $\beta = \beta' B$ mit $\beta' \in T^*$ und $B \in N$.

Analog : Bei der *linkslinearen Grammatik* wird das „$\beta = \beta' B$“ durch „$\beta = B\beta'$“ ersetzt.

Beispiel einer regulären Grammatik nach der Definition Y :

Man nehme an, zu der Sprache $L = \{a^*\} \cdot \{b^*\}$ existiert eine reguläre Grammatik $G = (N, T, P, \sigma)$, welche definierte ist mit $N = \{a, \sigma'\}$, $T = \{a, b\}$ und $P = \{\sigma \to a\sigma, \sigma \to \sigma', \sigma' \to \sigma' b, \sigma' \to \varepsilon\}$
$\Rightarrow L(G) = L$.

Satz Z :

Sei X ein Alphabet, dann existiert zu jeder regulären Sprache $L \subseteq X^*$ eine *rechtslineare Grammatik G*, führ die gilt: $L(G) = L$ und umgekehrt.

Beweis :

Der Beweis ist unterteilt in zwei Beweise für jeweils eine Richtung, also einer für „$\Rightarrow$“ und einer für „$\Leftarrow$“.

Beweis für „$\Rightarrow$“ :

Es sei die Sprache $L \subseteq X^*$ regulär. Dann existiert ein endlicher deterministischer Akzeptor $A = (X, S, \delta, s_0, F)$ mit der Eigenschaft $L = L(A)$.

Beweisidee :
Man konstruiere eine Grammatik aus der Zustandsüberführungsaktion δ.
Die Zustandsüberführungsaktion δ wird also zur Grammatik.

Beweisführung:
Es sei eine Grammatik $G = (N, T, P, \sigma)$ definiert durch:

$N := S$, $T := X$, $\sigma := s_0$ und $P := \{s \to xs' \mid \forall s, s' \in S : \delta(s, x) = s'\} \cup \{s \to \varepsilon \mid \forall s \in F\}$.

Nun ist zu zeigen: $L(G) = L$

Es sei ein Wort $w \in L(G)$ mit $w = x_1 \ldots x_k$, $x_i \in X$ $\underset{\text{definiert}}{\Leftrightarrow}$

$s_0 \xrightarrow[G]{} w \Leftrightarrow s_0 \to x_1 s_1 \to x_1 x_2 s_2 \to \ldots \to x_1 \ldots x_k \Leftrightarrow$

$\forall i \in \{0, \ldots, k-1\} : \delta(s_i, x_{i+1}) = s_{i+1}$ und $s_k \in F \Leftrightarrow \delta^*(s_0, w) = s_k \in F \Leftrightarrow w \in L(A) = L$.

Beweis für „$\Leftarrow$“ :
Es sei $G = (N, T, P, \sigma)$ eine rechtslineare Grammatik.

Beweisidee :
Wir konstruieren einen nichtdeterministischen Akzeptor $A = (X, S, \delta, s_0, F)$.

Beweisführung:
Es sei $A = (X, S, \delta, s_0, F)$ mit $L(G) = L(A)$ definiert durch :

$$S := N \cup \{\alpha\} \text{ mit } \alpha \notin N,\ s_0 := \{\sigma\} \text{und } F = \begin{cases} \{\sigma, \alpha\}, \text{ falls } \sigma \to \varepsilon \in \mathrm{P} \\ \{\alpha\}, \text{ falls } \sigma \to \varepsilon \notin \mathrm{P} \end{cases}.$$

Des weiteren gelten die folgenden Gesetzmäßigkeiten : $\begin{matrix} B \in \delta(A, x), \text{ falls } A \to xB \in P \\ \alpha \in \delta(A, x), \text{ falls } A \to x \in P \end{matrix}$.

Nun soll gelten für $n \geq 1$:

Ein Wort $w = x_1 ... x_n$ mit $x_i \in X$ ist Element von $L(G)$.

$\Leftrightarrow$ Es existiert eine Folge von *Nichtterminalsymbolen* $A_1, ..., A_{n-1}$ für die gilt:

$\sigma \to x_1 A_1 \to x_1 x_2 A_2 \to ... \to x_1 ... x_{n-1} A_{n-1} \to x_1 ... x_n = w$.

$\Leftrightarrow$ Es existiert eine Folge von Zuständen $A_1, ..., A_{n-1}$ mit

$A_1 \in \delta(\sigma, x_1), A_2 \in \delta(A_1, x_2), ..., \alpha \in \delta(A_{n-1}, x_n)$.

$\Leftrightarrow$ $w \in L(A)$. $\square$