Grundlagen der theoretischen Informatik

Kurt Sieber

Fakultät IV, Department ETI
Universität Siegen

SS 2013

Vorlesung vom 07.05.2013 (Teil 1)

Das Pumping Lemma (dt.: **Schleifensatz**) bietet eine weitere Methode, um zu zeigen, dass eine Sprache *nicht* regulär ist.

Es beschreibt eine *Eigenschaft* regulärer Sprachen. Um zu beweisen, dass eine Sprache *nicht* regulär ist, genügt es zu zeigen, dass sie diese Eigenschaft *nicht* besitzt.

Mit anderen Worten:
Das Pumping Lemma gibt eine *notwendige* Bedingung dafür an, dass eine Sprache regulär ist. Kann man zeigen, dass diese Bedingung nicht erfüllt ist, so weiß man, dass die Sprache nicht regulär ist.

Zum Vergleich:
Der Satz von Myhill und Nerode gibt eine *hinreichende und notwendige* Bedingung dafür an, dass eine Sprache regulär ist. Deshalb kann er für beide Zwecke benutzt werden: Sowohl zum Nachweis, dass eine Sprache regulär ist, als auch dazu, dass sie nicht regulär ist.

Die Idee:

Sei A ein DEA für die reguläre Sprache L, und sei n die Anzahl der Zustände von A.

Sei $x \in L$ mit $|x| \geq n$.

Dann werden $n+1$ Zustände bei der Abarbeitung von x auf A durchlaufen. Da A nur n Zustände hat, wird also mindestens ein Zustand mehrmals durchlaufen, d.h. der Weg, auf dem x abgearbeitet wird, enthält eine *Schleife*.

Durchläuft man diese Schleife öfter oder weniger oft als bei der Abarbeitung des Wortes x, so erhält man andere Wörter, die zum gleichen Endzustand führen wie x, die also ebenfalls in L liegen.

Wie sehen diese Wörter aus?

Sei $x = uvw$, wobei $v \neq \varepsilon$ das Teilwort ist, das in der Schleife abgearbeitet wird. Dann kann man v beliebig oft wiederholen, und erhält damit Wörter der Form $x_i = uv^iw$ ($i \geq 0$), die alle in L liegen.

Das Wort x_i entsteht sozusagen durch *Aufpumpen* des Wortes x an der Stelle v mit "Pumpfaktor" i. (Man beachte, dass $x_1 = uv^1w = x$, d.h. der Pumpfaktor 1 bewirkt *nichts*.)

Satz 2.35 (Pumping Lemma) *Sei $L \subseteq \Sigma^*$ regulär. Dann existiert eine Zahl $n \geq 1$, so dass für jedes $x \in L$ mit $|x| \geq n$ gilt: Es gibt eine Zerlegung $x = uvw$ mit $v \neq \varepsilon$ und $uv^iw \in L$ für alle $i \geq 0$.*

Intuition:
Wenn L regulär ist, dann kann man jedes hinreichend lange Wort $x \in L$ an mindestens einer Stelle v mit beliebigem Faktor i aufpumpen, ohne dass man die Sprache L verlässt.

Beweis:

Sei $A = (\Sigma, Q, s, F, \delta)$ ein DEA, der L erkennt.

Wir wählen $n = |Q|$.

Wenn $x \in L$ mit $|x| \geq n$, dann wird bei der Abarbeitung von x mindestens ein Zustand mehrmals besucht,

d.h. es existieren $p \in Q, q \in F$ und $u, v, w \in \Sigma^*$ mit $x = uvw$ und

$$(s, x) = (s, uvw) \vdash_A^* (p, vw) \vdash_A^+ (p, w) \vdash_A^* (q, \varepsilon)$$

Dabei ist $v \neq \varepsilon$, weil in $\vdash_A^+$ mindestens ein Zeichen verarbeitet wird.

Indem man die $\vdash_A^+$-Schleife i-mal wiederholt, folgt

$$(s, uv^i w) \vdash_A^* (p, v^i w) \vdash_A^+ \ldots \vdash_A^+ (p, w) \vdash_A^* (q, \varepsilon)$$

also $uv^i w \in L$ für jedes $i \geq 0$. $\square$

Korollar 2.36 *Zum Nachweis, dass L* nicht *regulär ist, genügt es zu zeigen: Für jedes $n \geq 1$ existiert ein $x \in L$ mit $|x| \geq n$, so dass für alle Zerlegungen $x = uvw$ mit $v \neq \varepsilon$ gilt: Es gibt ein $i \geq 0$ mit $uv^iw \notin L$.*

Intuition:
Es ist zu zeigen, dass *beliebig lange* Wörter $x \in L$ existieren, die sich an *jeder* Stelle v so aufpumpen lassen, dass das Ergebnis nicht mehr in L liegt.

Der Beweis, dass eine Sprache L nicht regulär ist, läuft also so ab: Für *jedes* $n \geq 1$ gibt man ein *passendes* Wort x mit $|x| \geq n$ an. Dann muss man für *jede* Zerlegung $x = uvw$ mit $v \neq \varepsilon$ einen *passenden* Pumpfaktor i wählen mit $uv^iw \notin L$.

Merke:
x und i darf man *passend wählen*, aber man muss *alle* Zerlegungen des Wortes x betrachten. Oft ist dabei eine Fallunterscheidung notwendig, weil man *alle* Möglichkeiten für die Position des Teilwortes v im Wort x durchspielen muss.

Beispiel:

$L_1 = \{a^n b^m \mid 0 \leq n \leq m\}$ ist nicht regulär, denn:

Sei $n \geq 1$.

Man wähle $x = a^n b^n$, also $|x| = 2n \geq n$.

Sei $x = uvw$ eine Zerlegung von x mit $v \neq \varepsilon$.

1. Fall: v liegt in a^n, d.h. $v = a^k$ für ein $k \geq 1$.

Dann gilt $uv^2w = a^{n+k}b^n \notin L_1$, weil $n + k > n$ wegen $k \geq 1$.

2. Fall: v liegt an der Schnittstelle, d.h. $v = a^i b^j$ mit $i, j \geq 1$.

Dann gilt $uv^2w = a^{n-i}a^ib^ja^ib^jb^{n-j} = a^nb^ja^ib^n \notin L_1$, denn wegen $i, j \geq 1$ stehen as hinter bs.

3. Fall: v liegt in b^m, d.h. $v = b^k$ für ein $k \geq 1$.

Dann gilt $uv^0w = a^nb^{n-k} \notin L_1$, weil $n - k < n$ wegen $k \geq 1$.

Am Beispiel wird klar:
Die Schwierigkeit bei der Anwendung des Pumping Lemmas besteht darin, dass man nicht weiß, wo das Teilwort v liegt. Das kann zu mühsamen Fallunterscheidungen führen oder ganz schiefgehen:

Beispiel:

$$L_2 = \{w \in \{a, b\}^* \mid \#_a(w) = \#_b(w)\}.$$

Dann gibt es für *jedes* $x \in L_2$ mit $|x| \geq 1$ eine Zerlegung $x = uvw$ mit $v \neq \varepsilon$, so dass $uv^iw \in L_2$ für *alle* $i \geq 0$.

Nämlich $u = w = \varepsilon$ und $v = x$.

Dann ist $uv^iw = x^i \in L_2$ für alle $i \geq 0$.

Also kann man gar kein passendes Wort x finden, um mit Satz 2.35 nachzuweisen, dass L_2 nicht regulär ist.

Mit anderen Worten:

L_2 erfüllt die im Pumping Lemma genannte Eigenschaft regulärer Sprachen, ohne selbst regulär zu sein. Also ist diese Eigenschaft eine notwendige, aber keine hinreichende Bedingung dafür, dass die Sprache regulär ist.

Die meisten Beispiele lassen sich bewältigen, wenn man das Pumping Lemma wie folgt verstärkt.

Satz 2.37 (Starkes Pumping Lemma) *Sei $L \subseteq \Sigma^*$ regulär. Dann existiert eine Zahl $n \geq 1$, so dass für jedes $x \in L$ mit $|x| \geq n$ gilt: Es gibt eine Zerlegung $x = uvw$ mit $v \neq \varepsilon$, $|uv| \leq n$ und $uv^iw \in L$ für alle $i \geq 0$.*

Man erhält diese Aussage, indem man den Beweis von Satz 2.35 leicht abändert: Anstelle des Wortes x betrachtet man das Präfix y von x der Länge n. Schon bei der Abarbeitung von y muss mindestens ein Zustand mehrmals auftauchen, also können wir das Wort v so wählen, dass es in y liegt, und das bedeutet $|uv| \leq n$.

Das starke Pumping Lemma eröffnet neue Möglichkeiten, weil wir für das (passend gewählte) Wort x nicht mehr *alle* Zerlegungen $x = uvw$ betrachten müssen, sondern nur noch diejenigen mit $|uv| \leq n$.

Damit vereinfacht sich z.B. der Beweis für $L_1 = \{a^n b^m \mid 0 \leq n \leq m\}$: Man wählt nach wie vor $x = a^n b^n$. Aber jetzt braucht man nur noch Zerlegungen der Form $x = uvw$ mit $|uv| \leq n$ zu untersuchen. $|uv| \leq n$ bedeutet aber, dass v ganz in a^n liegen muss, also bleibt von den drei Fällen im obigen Beweis nur der erste übrig.

Der Beweis für $L_2 = \{w \in \{a, b\}^* \mid \#_a(w) = \#_b(w)\}$ gelingt jetzt auch:

Sei $n \geq 1$.

Man wähle $x = a^n b^n$, also $x \in L_2$ und $|x| = 2n \geq n$.

Sei $x = uvw$ eine Zerlegung mit $v \neq \varepsilon$ und $|uv| \leq n$.

Dann liegt v in a^n, also $v = a^k$ für ein $k \geq 1$.

Also ist $uv^2w = a^{n+k}b^n \notin L_2$, weil $n + k \neq n$ wegen $k \geq 1$.

Weiteres Beispiel:

Sei $L_3 = \{a^p \mid p \text{ ist Primzahl}\}$.
Es wurde schon mit dem Satz von Myhill-Nerode bewiesen, dass L_3 nicht regulär ist. Mit dem Pumping Lemma geht es hier einfacher:

Sei $n \geq 1$.

Man wählt $x = a^p$ für eine Primzahl $p \geq n$, also ist $x \in L_3$ und $|x| \geq n$.
Sei $x = uvw$ eine Zerlegung mit $v \neq \varepsilon$, d.h. $v = a^k$ mit $k \geq 1$.
Dann gilt $uv^{p+1}w = a^{p+kp} = a^{p(1+k)} \notin L_3$, denn wegen $1 + k \geq 2$ ist p ein echter Teiler von $p(1+k)$, also $p(1+k)$ keine Primzahl.

Wie kommt man auf diesen Pumpfaktor?

Die Idee besteht darin, das Wort $x = a^p$ um ein Vielfaches von p zu *verlängern*. Mit Pumpfaktor p würde man das Teilwort $v = a^k$ durch $v^p = a^{kp}$ ersetzen, also x nur um $k(p-1)$ verlängern. Deshalb ist nicht p sondern $p+1$ der passende Pumpfaktor.

Reguläre Sprachen

Übrigens ist auch die im starken Pumping Lemma genannte Eigenschaft noch keine hinreichende Bedingung für die Regularität einer Sprache, aber in den meisten Fällen reicht es aus, um zu zeigen, dass eine Sprache nicht regulär ist.

Andere Verstärkungen des Pumping Lemmas sind denkbar, z.B. kann man $|uv| \leq n$ durch $|vw| \leq n$ ersetzen.

Offene Frage:

Kann man das Pumping Lemma so verstärken, dass man eine hinreichende und notwendige Bedingung für die Regularität der Sprache erhält?