Das Pumping Lemma der regulären Sprachen

Frank Heitmann
heitmann@informatik.uni-hamburg.de

1 Das Pumping Lemma

Das Pumping Lemma der regulären Sprachen macht eine Aussage der Art "*wenn* eine Sprache L regulär ist, *dann* besitzt L die Eigenschaft X", wobei die Eigenschaft X relativ kompliziert ist, aber dazu unten mehr. Wenn man nun also zeigt, dass eine gegebene Sprache die Eigenschaft X *nicht* besitzt, so darf man folgern, dass die Sprache nicht regulär ist. Wäre die Sprache nämlich regulär, so hätte sie ja (nach dem Pumping Lemma) gerade die Eigenschaft X!

Das Pumping Lemma der regulären Sprachen ist also ein "Werkzeug" mit dem man zeigen kann, das eine gegebene Sprache *nicht* regulär ist.

Ausformuliert lautet das Pumping Lemma wie folgt: Wenn eine Sprache L regulär ist, dann gibt es eine natürliche Zahl n derart, dass für alle Wörter $z \in L$ mit $|z| \geq n$ (also alle Wörter aus L deren Länge mindestens n ist) eine Zerlegung $z = uvw$ (also eine Zerlegung von z in drei Teilworte u, v und w, die aneinandergereiht z ergeben) existiert derart, dass die drei Eigenschaften $|uv| \leq n$, $|v| \geq 1$ und $uv^i w \in L$ ($i \geq 0$) erfüllt sind.[1]

Wenn eine Sprache L also regulär ist, dann gibt es eine natürliche Zahl n. Diese Zahl hängt von L ab, d.h. für verschiedene reguläre Sprachen, hat man hier im Allgemeinen auch verschiedene Zahlen. Ferner ist eine genaue Kenntnis dieser Zahl *nicht* nötig (und im Allgemeinen auch gar nicht gewünscht). Es genügt zu wissen, *dass* eine solche Zahl existiert. Das Pumping Lemma sagt weiter, dass nun jedes Wort z aus L (jedes $z \in L$), dessen Länge mindestens n ist ($|z| \geq n$) "zerlegt"[2] werden kann in $z = uvw$, wobei $|uv| \leq n$, $|v| \geq 1$

[1] Oft wird das Pumping Lemma formuliert als "Für alle regulären Sprachen gibt es eine natürliche Zahl...". Dies ist aber äquivalent zu obigen.

[2] Dieses "Zerlegen" macht bisweilen Probleme. Gemeint ist mit $z = uvw$ folgendes: z ist ein Wort z.B. *aabbacb*. Man kann nun die ersten i Buchstaben des Wortes nehmen und aus diesen das Wort u machen. Ist z.B. $i = 3$, so ist $u = aab$. Nun kann man von den *folgenden* Buchstaben z.B. die nächsten 2 nehmen und aus diesen das Wort v bilden, d.h. $v = ba$. Ebenso verfährt man mit w. Da uvw (also u ausgeschrieben, dann direkt dahinter v und dahinter w) das Wort z ergeben sollen, muss w bis zum letzten Buchstaben gehen (ebenso wie u mit dem ersten beginnen muss), d.h. es ist $w = cb$. $u = aab, v = ba, w = cb$ hintereinander geschrieben ergibt dann gerade *aabbacb* also z. Betrachten wir dann später Wörter wie z.B. $uv^2w = uvvw$, so erhält man dann weiter in obigen Beispiel das Wort *aabbabacb*. Die Teilwörter werden also wie Dominosteine aneinandergereiht. Man beachte zuletzt, dass die Teilwörter auch leer, also ϵ, sein dürfen. Z.B. ist für $z = aabbacb$ auch $u = \epsilon, v = aabb, w = acb$ eine erlaubte Zerlegung $z = uvw$.

und $uv^iw \in L$ für alle $i \geq 0$ gilt[3], d.h. die beiden Teilwörter u und v zusammen (aneinandergereiht) ergeben ein Wort (uv), dessen Länge höchstens n ist. Ferner hat das Teilwort v mindestens die Länge 1 (sprich: Das Teilwort v darf nicht das leere Wort sein, während u und w durchaus das leere Wort ϵ sein dürfen; siehe auch die vorherige Fussnote zur Zerlegung) und zuletzt gilt $uv^iw \in L$ für alle $i \geq 0$, d.h. es ist nicht nur $z \in L$ und damit wegen $z = uvw$ auch $uvw \in L$ (z und uvw sind ja das gleiche Wort und nur verschieden notiert), sondern es sind auch die Wörter $uv^0w = uw$, $uv^1w = uvw$, $uv^2w = uvvw$, $uv^3w = uvvvw$ usw. alle in der Menge L. Man spricht davon, dass man an dem v "pumpen" kann, da man das Teilwort v beliebig oft notieren kann.[4]

Dies ist zunächst die Aussage des Pumping Lemmas. Im nächsten Abschnitt benutzen wir das Pumping Lemma als Werkzeug um nachzuweisen, dass bestimmte Mengen nicht regulär sind.

2 Das Pumping Lemma als Werkzeug

Wir wollen nun das Pumping Lemma benutzen, um nachzuweisen, dass eine gegebene Sprache nicht regulär ist.[5] Oben wurde schon darauf hingewiesen, dass man hierzu das Pumping Lemma benutzen kann und *nicht* dazu z.B. zu zeigen, dass eine Sprache regulär ist. Dazu müsste die Aussage nämlich sein, dass jede Sprache, die die Eigenschaft X besitzt, regulär ist - und das ist nicht die Aussage des Pumping Lemmas, sondern nur, dass wenn eine Sprache regulär ist, sie die Eigenschaft X besitzt.

Wie wird nun das Pumping Lemma benutzt, um zu zeigen, dass eine gegebene Sprache nicht regulär ist? Wir illustrieren dies an einem Beispiel. Sei

$$L = \{a^nb^n \mid n \geq 0\}$$

Die Sprache L enthält neben dem leeren Wort (bei $n = 0$ ist $a^nb^n = a^0b^0 = \epsilon\epsilon = \epsilon$) die Wörter ab, $aabb$, $aaabbb$, $aaaabbbb$ usw. Diese Sprache ist nicht regulär, wie man mit dem Pumping Lemma nachweisen kann.

Als erstes nimmt man dazu an, die Sprache wäre regulär.[6] Wie schon mehrfach geschrieben, ist die Aussage des Pumping Lemmas, wenn eine Sprache regulär ist, dann gilt X. Da wir L als regulär annehmen, muss X (für L) also gelten. Dies wird nun zum Widerspruch geführt indem gezeigt wird, dass L die Eigenschaft X nicht besitzt. Die Aussage X ist, dass eine natürliche Zahl n existiert mit bestimmten Eigenschaften. Wir müssen also zeigen, dass keine solche

[3]Man beachte nochmal, dass die Aussage ist: Für jedes $z \in L$ mit $|z| \geq n$ *existiert* eine Zerlegung in $z = uvw$ mit den drei Eigenschaften. Wenn wir nachher also zeigen wollen, dass eine Sprache diese Eigenschaft nicht hat, dann werden wir zeigen müssen, dass *keine* solche Zerlegung existiert, dass also *alle* Zerlegungen von z in uvw im Widerspruch zu einer der drei Eigenschaften stehen. Dazu gleich mehr.

[4]Aber stets an der Stelle, wo das ursprüngliche v auch stand, d.h. es ist nicht erlaubt das Wort $uvwv$ zu bilden. Die v's stehen immer direkt beieinander.

[5]Regulär ist eine Sprache L genau dann, wenn man einen endlichen Automaten A angeben kann, der die Sprache akzeptiert, für den also $L(A) = L$ gilt.

[6]Das ganze wird also ein Widerspruchsbeweis. Wir nehmen an die Sprache L sei regulär und zeigen dann, dass dies zu einem Widerspruch führt.

Zahl existiert, d.h. das für alle natürlichen Zahlen die geforderten Eigenschaften nicht erfüllt sind. Dies macht man, indem man im Prinzip einen zweiten Widerspruchsbeweis macht. Man nimmt nun nämlich an, dass diese Zahl existiert und gibt ihr einen Namen, z.B. k. Da k nicht weiter spezifiziert ist, hat man die Zahl beliebig gewählt. Zeigt man nun, dass k die gewünschten Eigenschaften nicht erfüllt, so hat man gezeigt, dass keine Zahl dies tut. Im Beweis tut man dies, indem man zunächst annimmt, dass L regulär ist, dann folgert, dass also das Pumping Lemma gelten muss (insb. die Konklusion des Pumping Lemmas, nämlich dass eine natürliche Zahl existiert mit...) und das k nun die (natürliche) Zahl aus dem Pumping Lemma sei. Die Aussage ist nun weiter, dass sich jedes Wort $z \in L$ mit $|z| \geq k$ zerlegen lässt in $z = uvw$ (also eine Zerlegung von z in uvw existiert) mit den drei Eigenschaften. Um zu zeigen, dass dies nicht gilt (also um einen Widerspruch zu erzeugen), muss man nun zeigen, dass *ein* Wort $z \in L$ mit $|z| \geq k$ existiert derart, dass *alle* Zerlegungen von z in uvw mindestens eine der drei Eigenschaften nicht erfüllen.[7]

Man muss nun also ein Wort z wählen, dessen Länge größer ist als k. Da man nichts genauer über k weiss, muss z abhängig von k gewählt werden. In obigen Beispiel kann man z.B. $z = a^k b^k$ wählen. Andere Möglichkeiten wären $z = a^{k/2} b^{k/2}$ oder $z = a^{2k} b^{2k}$. Man beachte, dass der "Exponenten" des a's und des b's immer gleich ist, da sonst $z \in L$ nicht erfüllt ist. Eine schlechte Wahl für z wäre $z = a^3 b^3$ oder auch $z = a^{k/3} b^{k/3}$. Im letzteren Fall ist die Länge des Wortes kleiner als k (nämlich nur $2 \cdot \frac{k}{3}$) und im ersteren Fall ist nicht klar, ob die Länge grösser ist als k, da man nicht weiss, wie k genau aussieht (hier ist das Problem also, dass z unabhängig von k gewählt wurde). Wir betrachten nachfolgend das Wort $z = a^k b^k$ und zeigen dann, dass jede Zerlegung von z in uvw im Widerspruch zu einer der drei Bedingungen

- $|uv| \leq k$
- $|v| \geq 1$
- $uv^i w \in L$ für alle $i \geq 0$

steht.[8]

Um dies zu erreichen geht man wie folgt vor: Man betrachtet nur Zerlegungen von z in uvw, die die erste und zweite Bedingungen erfüllen. Jene Zerlegungen nämlich die dies nicht tun, erfüllen ja bereits mindestens eine der drei Bedingungen nicht (nämlich die erste, die zweite oder sogar beide). Von diesen Zerlegungen zeigt man dann, dass sie die dritte Bedingung nicht erfüllen und hat so gezeigt, dass jede Zerlegung mindestens eine der drei Bedingungen

[7] Man beachte, wie hier im Prinzip die in der Logik bereits aufgetretenden Quantoren und ihre Negation auftritt. Die ursprüngliche Aussage ist, es existiert eine Zerlegung mit $A \wedge B \wedge C$ (für die drei Eigenschaften). Wollen wir dies Widerlegen, also die Negation davon zeigen, so wird daraus, dass alle Zerlegungen (aus dem Existenzquantor wird ein Allquantor) $\neg A \vee \neg B \vee \neg C$ erfüllen, also mindestens eine der drei Eigenschaften nicht erfüllen.

[8] Nocheinmal zur Betonung: Es genügt dies für *ein* Wort zu zeigen. Man muss aber für *ein* Wort zeigen, dass für *alle* Zerlegungen in uvw mindestens *eine* der obigen drei Bedingungen nicht erfüllt ist.

nicht erfüllt.[9] Von den Zerlegungen, die die erste und zweite Bedingung erfüllen gibt es im Allgemeinen (und auch hier) mehr als eine. Alle diese Zerlegungen haben jedoch (meist) bestimmte Eigenschaften und können zum Widerspruch zur dritten Eigenschaft geführt werden. Hier geht dies wie folgt: Eine Zerlegung $z = a^k b^k = uvw$, die die erste Bedingung erfüllt, für die also $|uv| \leq k$ gilt, muss $uv \in \{a\}^*$ erfüllen, denn uv steht am Anfang von z und kann nicht mehr als die ersten k Buchstaben von z einnehmen, muss also nur aus a's bestehen. Die zweite Bedingung, dass $|v| \geq 1$ gilt, führt dann zu $uv \in \{a\}^+$, bzw. $v \in \{a\}^+$, da v aus mindestens einem Buchstaben bestehen muss. Es gibt also ein j mit $1 \leq j \leq k$ mit $a^j = v$. $j \geq 1$, da v ja aus mindestens einem Buchstaben besteht und $j \leq k$, da ja $|uv| \leq k$ gelten muss, also insb. auch $|v| \leq k$. Die einzelnen Zerlegungen von z in uvw, die die erste und zweite Bedingung erfüllen unterscheiden sich nun dadurch, welchen Wert j genau hat und wo v beginnt (d.h. bis wohin u geht). Besteht u aus den ersten p a's (wobei $p = 0$ erlaubt ist), so ist also $u = a^p$, $v = a^j$ und $w = a^{k-p-j} b^k$, d.h. wir haben $z = uvw = a^p a^j a^{k-p-j} b^k$, wobei $p + j \leq k$ und $j \geq 1$ gilt. Jede dieser Zerlegungen (nochmal: Je nachdem, welche Werte man genau für p und j wählt erhält man unterschiedliche Zerlegungen) steht nun im Widerspruch zu der dritten Bedingung, die besagt, dass $uv^i w \in L$ für alle $i \geq 0$ gelten muss. Wir betrachten das Wort $uv^2w = uvvw$. Dieses Wort ist dann $a^p a^j a^j a^{k-p-j} b^k = a^{k+j} b^k$ und wegen $j \geq 1$ hat dieses Wort nicht mehr die für Worte in L benötigte Form (das auf eine Anzahl von a's die gleiche Anzahl von b's folgen).

Damit ist der Beweis abgeschlossen. Es wurden alle Zerlegungen betrachtet, die die erste und zweite Bedingung erfüllen und gezeigt, dass diese Zerlegungen die dritte Bedingung nicht erfüllen. Dies wurde für die (nicht näher spezifizierte) Zahl aus dem Pumping Lemma gezeigt, die existieren müsste, wenn die Sprache regulär wäre. Folglich ist die Sprache nicht regulär.

Im folgenden beweisen wir für eine weitere Sprache, das sie nicht regulär ist, gehen im Beweis aber schneller und mit weniger ausführlichen Erläuterungen vor.

Sei $L = \{a^n b^m \mid 0 \leq m \leq n\}$. Wir wollen zeigen, dass L nicht regulär ist und nehmen dazu an L wäre regulär. Es müsste dann das Pumping Lemma gelten. Sei k die Zahl aus dem Pumping Lemma. Wir betrachten das Wort $z = a^k b^k \in L$. Für jede Zerlegung $z = uvw$, die die erste und zweite Bedingung des Pumping Lemmas erfüllt, gilt $uv \in \{a\}^+$ und $v = a^j$ für ein $j \geq 1$. Betrachten wir nun das Wort $uv^0w = uw = a^{k-j} b^k$, so gilt $uw \notin L$, da in uw auf $k - j < k$ a's k viele b's folgen (also mehr als zuvor a's in dem Wort waren). $uw \notin L$ steht aber im Widerspruch zur dritten Bedingung des Pumping Lemmas. Daher ist die ursprüngliche Annahme falsch und L als nicht regulär nachgewiesen.

Zuletzt sei darauf hingewiesen, dass die Wahl des Wortes z ein kreativer Akt ist (in dem Sinne, dass es dafür kein "Rezept" gibt) und nicht immer einfach ist. Es gibt nicht-reguläre Sprachen, bei denen man sowohl Worte finden kann, bei denen die obigen Schritte zum Erfolg führen, als auch Worte, die zwar länger

[9] In einer vorherigen Fussnote wurde bereits gesagt, dass $\neg A \vee \neg B \vee \neg C$ zu zeigen ist (A, B, C stehen dabei jeweils für eine der drei Bedingungen). Dies ist äquivalent zu $\neg(A \wedge B) \vee \neg C$, was wiederum äquivalent ist zu $(A \wedge B) \Rightarrow \neg C$ und dies ist gerade das, was gezeigt wird.

sind als die (angenommene) Zahl aus dem Pumping Lemma, bei denen aber nicht jede Zerlegung im Widerspruch zu einer der drei Bedingungen steht. Bei solchen Sprachen muss man also u.U. mehrere Ansätze für das zu wählende Wort ausprobieren. Ferner gibt es auch nicht-reguläre Sprachen, die sich mit dem Pumping Lemma nicht als nicht-regulär nachweisen lassen, die also die Eigenschaften des Pumping Lemmas erfüllen, aber dennoch nicht regulär sind (darum ist das Pumping Lemma auch nicht dafür zu benutzen eine Sprache als regulär nachzuweisen, sondern nur umgekehrt, um nachzuweisen, dass eine Sprache nicht regulär ist).

3 Zur Übung

Zur Übung können zunächst die obigen beiden Sprachen $L = \{a^n b^n \mid n \geq 0\}$ und $L' = \{a^n b^m \mid 0 \leq m \leq n\}$ genommen werden und ohne diesen Text nocheinmal versucht werden diese als nicht-regulär nachzuweisen.

Zur weiteren Übung können die folgenden beiden Sprachen dienen, wobei die Frage ist, *ob* sie regulär sind oder nicht. Eine davon ist tatsächlich regulär (Probieren Sie dafür einen Automaten zu konstruieren!), die andere ist es nicht:

- $\{a^n b^m \mid \exists g \in \mathbb{N} : 2g = n + m\}$
- $\{a^n b^m \mid \exists g \in \mathbb{N} : 2g = n = m\}$

Ein Versuch einen Automaten zu konstruieren kann dabei dazu dienen, ein Gefühl zu entwickeln, warum die Sprache vielleicht doch nicht regulär ist. Dann kann man mit dem Pumping Lemma versuchen dies nachzuweisen. Andersherum sollte man, wenn man mit dem Pumping Lemma nicht vorankommt nachzuweisen, dass eine Sprache nicht-regulär ist, versuchen einen Automaten zu entwickeln, da die Sprache dann vielleicht doch regulär ist. (Schlägt dies fehl, so braucht man aber vielleicht lediglich andere Mittel, um die Sprache als nicht-regulär nachzuweisen, als das Pumping Lemma.)

Die Benutzung des Pumping Lemma erfordert einiges an Übung also nicht gleich entmutigen lassen, wenn es nicht sofort klappt!