Automaten und Formale Sprachen

Pumping-Lemma für reguläre Sprachen

Zerlegen komplizierterer Wörter

Lara Stoltenow

27. Mai 2019

Inhaltsverzeichnis

Dieses Dokument soll einen kurzen Überblick darüber geben, wie man beim Pumping-Lemma systematisch Wörter zerlegt, ohne dabei Zerlegungen zu vergessen.

Es ist als Ergänzung zu Tutorium und Übung zu sehen. Es ersetzt *nicht* die Vorlesung.

Für dieses Dokument gilt: es geht nur um die Zerlegungen. Da die Zerlegungen nur vom gewählten Wort x abhängen, geben wir zu den Beispielen auch keine Sprache an. (Man kann sich allerdings leicht Beispielsprachen daraus erstellen.)

1 Ein einfaches Beispiel

Wir beginnen mit einem einfachen Beispiel und zerlegen das Wort

$$x = a^n b^n$$

Wir überlegen uns zunächst, welche Einschränkungen die Zerlegungen haben. Es soll $x = uvw$ gelten mit $|uv| \leq n$ und $|v| \geq 1$.

Insbesondere die erste Einschränkung spart uns viel Arbeit beim Zerlegen. Wir schauen dazu die ersten n Zeichen des gewählten Wortes an. Die Kombination von u und v kann nicht mehr als diese Zeichen abdecken (da sonst $|uv| > n$), also müssen wir auch nur innerhalb der ersten n Zeichen verschiedene Aufteilungsmöglichkeiten suchen.

Für das aktuelle Beispiel gilt also: die ersten n Zeichen sind a^n, d.h. sowohl u als auch v können nur aus as bestehen. Alle bs (und ggf. überzählige as, falls $|uv| < n$) befinden sich daher in w. Die genaue Zahl der as und bs wissen wir zwar nicht, aber wir können diese unproblematisch mit Variablen ausdrücken und erhalten folgende Zerlegung:

$$u = a^k,\ v = a^\ell,\ w = a^{n-k-\ell} b^n$$

Da dies bereits alle Möglichkeiten abdeckt, wie man die as auf u und v verteilen kann (wir haben keine zusätzlichen Einschränkungen notiert, wobei wir aber wegen $|v| \geq 1$ wenigstens $\ell \geq 1$ hinzufügen sollten), benötigen wir keine weiteren Zerlegungen.[1]

[1] Manche Studenten geben eine zweite Zerlegung $u = \varepsilon,\ v = a^\ell,\ w = a^{n-\ell} b^n$ an. Das ist nicht falsch, aber diese ist mit $k = 0$ bereits in der zuvor angegebenen Zerlegung enthalten, d.h. man müsste, wenn man beide Fälle angibt, später beim Pumpen auch wirklich beide Fälle pumpen (also zusätzliche Arbeit).

2 Wiederholte Fallunterscheidung, einführendes Beispiel

Manchmal steht man vor dem Problem, dass das zu zerlegende Wort scheinbar viele Möglichkeiten der Zerlegung bietet. Es erscheint nicht offensichtlich, welche Zerlegungen es überhaupt gibt.

In solchen Fällen kann es sinnvoll sein, die Zerlegungen zunächst anhand eines einfachen Kriteriums in zwei Gruppen zu unterteilen. (Bei dieser Unterteilung muss man sich nicht notwendigerweise als Ziel setzen, eine brauchbare Zerlegung zu erhalten.) Falls sich daraus dann eine eindeutige Zerlegung gibt, schreibt man sie auf, andernfalls unterteilt man solange weiter, bis das möglich ist.

Als Beispiel sei folgendes Wort gegeben:

$$x = abc^n d^n$$

Es erscheint nicht offensichtlich, wie dieses Wort (vollständig) zerlegt werden kann. Jedoch können wir sicher sein, dass eine der folgenden Möglichkeiten zutrifft:

1. u enthält das a
2. u enthält das a nicht

Diese beiden Möglichkeiten sind prinzipiell erst einmal denkbar (es spielt keine Rolle, ob diese tatsächlich zu gültigen Zerlegungen führen) und decken alle Möglichkeiten ab (entweder enthält u das a oder nicht, es gibt keine anderen Möglichkeiten).

Dies hilft uns aber noch nicht dabei, die Zerlegungen auch tatsächlich aufzuschreiben ($u = \dots, v = \dots, w = \dots$). Darum unterteilen wir den ersten Teil weiter:

1. u enthält das a und ...
   a) ... u enthält das b
   b) ... u enthält das b nicht
2. u enthält das a nicht

Auch hier decken die neuen Fälle wieder alle denkbaren Möglichkeiten ab.

Wir stellen fest, dass die beiden ersten Fälle sich nun recht leicht zu vollständigen Zerlegungen ergänzen lassen: im ersten Fall enthält also u sowohl das a als auch das b (und möglicherweise cs). Wegen $|uv| \leq n$ kann dann auch v nur cs enthalten. Die Zahl der cs in uv kann höchstens $n-2$ sein (da es maximal n Zeichen sein können und zwei davon bereits durch das ab verbraucht sind). Analog entsteht der zweite Unterfall. Wir erhalten:

1. u enthält das a und ...
   a) ... u enthält das b:
      $u = abc^k,\ v = c^\ell,\ w = c^{n-k-\ell}d^n$ (wobei $k \geq 0,\ \ell \geq 1,\ k+\ell \leq n-2$)
   b) ... u enthält das b nicht:
      $u = a,\ v = bc^k,\ w = c^{n-k}d^n$ (wobei $0 \leq k \leq n-2$)
2. u enthält das a nicht

Wir können nun den zweiten Punkt ergänzen: wenn u das a nicht enthält, muss u leer sein. Da v nicht leer sein darf, muss v mindestens dieses a enthalten.[2] Es ist allerdings nicht klar, ob v auch das nachfolgende b enthält, d.h. wir führen weitere Unterfälle ein:

1. u enthält das a und ...
   a) ... u enthält das b:
      $u = abc^k,\ v = c^\ell,\ w = c^{n-k-\ell}d^n$ (wobei $k \geq 0,\ \ell \geq 1,\ k+\ell \leq n-2$)
   b) ... u enthält das b nicht:
      $u = a,\ v = bc^k,\ w = c^{n-k}d^n$ (wobei $0 \leq k \leq n-2$)

[2] Man könnte hier durchaus eine weitere Fallunterscheidung „v enthält a“ oder „v enthält kein a“ durchführen. Hiervon führt dann nur der erste Fall zu weiteren Zerlegungen, während der zweite Fall nur $u = v = \varepsilon$ erzeugt, was wegen $|v| \geq 1$ nicht möglich ist.

2. u enthält das a nicht und ...
   a) ... v enthält das b
   b) ... v enthält das b nicht

Damit können wir jetzt die Zerlegungen komplett angeben:

1. u enthält das a und ...
   a) ... u enthält das b:
   $u = abc^k,\ v = c^\ell,\ w = c^{n-k-\ell}d^n$ (wobei $k \geq 0,\ \ell \geq 1,\ k + \ell \leq n - 2$)
   b) ... u enthält das b nicht:
   $u = a,\ v = bc^k,\ w = c^{n-k}d^n$ (wobei $0 \leq k \leq n - 2$)
2. u enthält das a nicht und ...
   a) ... v enthält das b:
   $u = \varepsilon,\ v = abc^k,\ w = c^{n-k}d^n$ (wobei $0 \leq k \leq n - 2$)
   b) ... v enthält das b nicht:
   $u = \varepsilon,\ v = a,\ w = bc^nd^n$

Damit haben wir alle vier Zerlegungen dieses Wortes angegeben. Bei den Zwischenschritten haben wir dabei nur Entscheidungen getroffen, von denen wir uns sicher waren, dass alle Möglichkeiten abgedeckt sind (enthält oder enthält nicht).

3 Unterteilung in mehrere Fälle

Je nach Komplexität der betrachteten Sprache kann man auch mehr als zwei Fälle benötigen. Wir zeigen das am Beispiel des Wortes

$$x = (ab)^n c^n$$

Eine Möglichkeit ist, im ersten Schritt folgende Fälle zu schaffen:

1. u endet mit a
2. u endet mit b
3. u ist leer

Man beachte den dritten Fall, der notwendig ist, um wirklich alle denkbaren Möglichkeiten abzudecken.

Für diese Fälle können wir jeweils schon angeben, welche Form u haben muss:

1. u endet mit a ($u = (ab)^k a$ mit $k \geq 0$) und ...
2. u endet mit b ($u = (ab)^k$ mit $k \geq 1$) und ...
3. u ist leer ($u = \varepsilon$) und ...

Für den ersten Fall etwa wissen wir, dass v mit einem b beginnt und möglicherweise abs folgen. Was wir nicht wissen, ist, ob eine solche ab-Sequenz in der Mitte geteilt wird. Ähnliches gilt für die anderen beiden Fälle. Wir unterteilen also weiter:

1. u endet mit a ($u = (ab)^k a$ mit $k \geq 0$) und ...
   a) ... v endet mit b ($v = b(ab)^\ell$ mit $\ell \geq 0$)
   b) ... v endet mit a ($v = b(ab)^\ell a$ mit $\ell \geq 0$)
2. u endet mit b ($u = (ab)^k$ mit $k \geq 1$) und ...
   a) ... v endet mit b ($v = (ab)^\ell$ mit $\ell \geq 1$)
   b) ... v endet mit a ($v = (ab)^\ell a$ mit $\ell \geq 0$)
3. u ist leer ($u = \varepsilon$) und ...
   a) ... v endet mit b ($v = (ab)^\ell$ mit $\ell \geq 1$)
   b) ... v endet mit a ($v = (ab)^\ell a$ mit $\ell \geq 0$)

Die unterschiedlichen Einschränkungen für ℓ ergeben sich daraus, dass $|v| \geq 1$ teilweise durch fest vorgegebene Zeichen sichergestellt wird (wie etwa in Fall 1a das b am Anfang) und teilweise nicht. Wir vervollständigen die sechs Fälle nun zu kompletten Zerlegungen:

1. u endet mit a ($u = (ab)^k a$ mit $k \geq 0$) und ...
   a) ... v endet mit b:
      $u = (ab)^k a,\ v = b(ab)^\ell,\ w = (ab)^{n-k-\ell-1} c^n$ mit $\ell \geq 0$
   b) ... v endet mit a:
      $u = (ab)^k a,\ v = b(ab)^\ell a,\ w = b(ab)^{n-k-\ell-2} c^n$ mit $\ell \geq 0$
2. u endet mit b ($u = (ab)^k$ mit $k \geq 1$) und ...
   a) ... v endet mit b:
      $u = (ab)^k,\ v = (ab)^\ell, w = (ab)^{n-k\ell} c^n$ mit $k \geq 1,\ \ell \geq 1$
   b) ... v endet mit a:
      $u = (ab)^k,\ v = (ab)^\ell a,\ w = b(ab)^{n-k-\ell-1}$ mit $k \geq 1,\ \ell \geq 0$
3. u ist leer ($u = \varepsilon$) und ...
   a) ... v endet mit b:
      $u = \varepsilon,\ v = (ab)^\ell,\ w = (ab)^{n-\ell} c^n$ mit $\ell \geq 1$)
   b) ... v endet mit a:
      $u = \varepsilon,\ v = (ab)^\ell a,\ w = b(ab)^{n-\ell-1} c^n$ mit $\ell \geq 0$)

Dass bei dem Restwort in w manchmal noch 1 oder 2 abgezogen wird, liegt an aufgeteilten abs: in Fall 2b beispielsweise ist das a eines abs in v und das b in w, d.h. es gibt ein ab, welches nicht aus $(ab)^k$ oder $(ab)^\ell$ entstanden ist und darum zusätzlich zu k und ℓ abgezogen werden muss.

Dies ist nur eine Möglichkeit, um dieses Wort vollständig zu zerlegen. Man kann sich auch etwas Arbeit sparen und im ersten Schritt die Fälle „v beginnt mit b“ und „v beginnt mit a schaffen“ (der Fall „v ist leer“ kann wegen $|v| \geq 1$ nicht vorkommen!), wodurch oben die Fälle 2 und 3 zu einem einzigen Fall ($u = (ab)^k$ mit $k \geq 0$) zusammengefasst werden können. Die Zerlegung von oben in insgesamt sechs Fälle ist aber genauso richtig (es müssten später beim Pumpen nur zwei zusätzliche Fälle betrachtet werden).