Formale Methoden 1

Gerhard Jäger
Gerhard.Jaeger@uni-bielefeld.de

Uni Bielefeld, WS 2007/2008

23. Januar 2008

Das Pumping-Lemma

- Sein L eine **unendliche** reguläre Sprache über ein endliches Alphabet Σ.
- Es gibt einen NFA A, der L erkennt.
- Es gibt eine Zahl n, so dass A n Zustände hat.
- Fast alle Wörter in L bestehen aus mehr als n Buchstaben.
  - Sei $x \in L$, mit $l(x) > n$.
  - Es gibt einen Lauf von A, der x erkennt.
  - Weil A n Zustände hat und $l(x) > n$, wird mindestens ein Zustand von A mehrfach besucht. Sei Z derjenige Zustand, der als erster mehrfach besucht wird.
  - x lässt sich als $y \frown z \frown w$ darstellen, so dass
    - zwischen Startzustand und Z wird y erkannt,
    - zwischen dem ersten und dem zweiten Besuch von Z wird z erkannt, und
    - zwischen dem zweiten Besuch von Z und dem Endzustand wird w erkannt.

Das Pumping-Lemma

- Also:
  - die Schleife von Z nach Z, während der x erkannt wird, kann beliebig oft wiederholt werden.
- Also: $x \frown y^i \frown z \in L$, für beliebige $i \geq 0$.

Das Pumping-Lemma

Diese Überlegungen gelten für beliebige unendliche reguläre Sprachen.

Theorem

Sei L eine unendliche reguläre Sprache über das Alphabet Σ. Dann gibt es eine Zahl n, so dass sich alle Wörter $x \in L$ mit $l(x) \geq n$ zerlegen lassen in $x = y \frown z \frown w$, so dass folgende Eigenschaften erfüllt sind:

1. $l(z) \geq 1$,
2. $l(y) + l(z) \leq n$, *und*
3. *für alle $i \in \mathbb{N}$ gilt:* $y \frown z^i \frown w \in L$.

Anwendungen des Pumping-Lemmas

Das Pumping-Lemma ist nützlich, wenn man beweisen will, dass eine bestimmte Sprache **nicht** regulär ist.

- Beispiel: $L = \{a^m b^m | m > 0\}$ ist nicht regulär.
- Beweis:
  - Angenommen, L wäre regulär.
  - Dann gibt es ein n mit der im Pumping-Lemma genannten Eigenschaft (die Anzahl der Zustände des Automaten, der L erkennt).
  - $a^n b^n \in L$.
  - $a^n b^n = x \frown y \frown z$, mit $l(x \frown y) \leq n$, $l(y) \geq 1$, und $x \frown z \in L$.
  - $y = a^j$, für ein $j \geq 1$.
  - Also $x \frown z = a^{n-j} b^n \in L$, was einen Widerspruch zur Definition von L darstellt.
  - Also ist L nicht regulär.

Anwendungen des Pumping-Lemmas

- $L = \{a^n b^m | m \geq n\}$ ist nicht regulär.
- Beispiel: $L = \{a^n b^m | m > n > 0\}$ ist nicht regulär.
- Beweis:
  - Angenommen, L wäre regulär.
  - Dann gibt es ein $n > 0$ mit der im Pumping-Lemma genannten Eigenschaft.
  - $a^n b^n \in L$.
  - $a^n b^n = x \frown y \frown z$, mit $l(x \frown y) \leq n$, $l(y) \geq 1$, und $x \frown y^z \in L$.
  - $y = a^j$, für ein $j \geq 1$.
  - Also $x \frown y^{n \cdot m} \frown z \in L$, was einen Widerspruch zur Definition von L darstellt.
  - Also ist L nicht regulär.

Anwendungen des Pumping-Lemmas

- Auf ähnliche Weise lässt sich zeigen, dass die folgenden Sprachen nicht regulär sind:
  - $\{w \frown w | w \in \Sigma^*\}$ (die „Kopiersprache")
  - $\{w \frown w^R | w \in \Sigma^*\}$ (die „Spiegel-Sprache" oder „Palindrom-Sprache")
- Komplizierter:

$$L = \{x \in \{a,b\}^* | \text{Anzahl der } a \text{ in } x = \text{Anzahl der } b \text{ in } x\}$$

Anwendungen des Pumping-Lemmas

Zum Beweis, dass L nicht regulär ist, benötigt man die folgende Einsicht:

Theorem

Wenn L_1 und L_2 regulär sind, dann ist auch $L_1 \cap L_2$ regulär.

Zunächst ist zu zeigen, dass das Komplement einer regulären Sprache auch regulär ist. Das ist nahezu offensichtlich: wenn ein DFA A die Sprache L akzeptiert., dann muss man lediglich die Endzustände in A in Nicht-Endzustände und die Nicht-Endzustände in Endzustände umwandeln, um einen DFA zu erhalten, der $\Sigma^* - L$ akzeptiert. In der letzten Vorlesung hatten wir gesehen, dass die Vereinigung zweier regulärer Sprachen auch regulär ist.
Wenn L_1 und L_2 also regulär sind, dann sind das auch $\overline{L_1}$ und $\overline{L_2}$, also auch $\overline{L_1} \cap \overline{L_2}$, also auch $\overline{\overline{L_1} \cap \overline{L_2}}$, was nach dem de-Morganschen Gesetz gleich $L_1 \cap L_2$ ist.

Anwendungen des Pumping-Lemmas

- Beweis, dass
  $L = \{x \in \{a, b\}^* |$Anzahl der a in $x =$ Anzahl der b in $x\}$
  nicht regulär ist:
  - a^*b^* ist regulär, weil diese Sprache durch einen regulären Ausdruck beschrieben wird.
  - Angenommen, L ist regulär. Dann muss auch
    $L \cap a^*b^* = \{a^n b^n | n \geq 0\}$ regulär sein.
  - Oben wurde gezeigt, dass diese Sprache nicht regulär ist. Also ist auch L nicht regulär.

Ist Deutsch regulär?

Mit Hilfe des Pumping-Lemmas lässt sich zeigen, dass natürliche Sprachen nicht regulär sind. Ein mögliches Argument fürs Deutsche geht folgendermaßen:

- Mit Hilfe von „entweder ... oder ...“ kann man im Deutschen beliebig lange Sätze konstruieren:

Entweder es regnet, oder es schneit.

Entweder Hans glaubt, dass es entweder regnet oder dass es hagelt, oder es schneit.

Entweder Hans glaubt, dass es entweder so aussieht, als ob es entweder kalt regnet oder warm regnet, oder dass es hagelt, oder es schneit.

...

- Insbesondere gibt es eine unendlich lange Liste S_i von Sätzen wachsender Länge, so dass S_i genau i mal *entweder* enthält, und in denen alle *entweder*s allen *oder*s vorangehen.

Ist Deutsch regulär?

- Zu jedem *entweder* in einem deutschen Satz gehört ein *oder*; die Anzahl der *oder*s ist also mindestens so groß wie die Zahl der *entweder*s.
- Reguläre Sprachen sind unter Tilgung von einzelnen Elementen von Σ abgeschlossen: Wenn ich einen bestimmten Buchstaben, z.B. a, in jedem Wort einer regulären Sprache L tilge, erhalte ich wieder eine reguläre Sprache. (Beweis: Ersetze in dem regulären Ausdruck, der L beschreibt, alle Vorkommen von a durch ϵ.)
- D' enthält die undendliche Sequenz S_i', das Resultat der Tilgung aller Morphem außer *entweder* und *oder* in S_i

Ist Deutsch regulär?

- Angenommen, Deutsch ist regulär. Genauer gesagt würde das bedeuten, dass die Menge D der grammatischen deutschen Sätze eine reguläre Sprache ist über das Alphabet $\Sigma =$ die Menge aller deutschen Morpheme.
- Dann ist auch die Sprache D' regulär, die man erhält, wenn man in allen deutschen Sätzen alle Morpheme außer *entweder* und *oder* tilgt.
- Sei n die Zahl der Zustände des Automaten, der D' akzeptiert.
- $S'_n = \text{entweder}^n\text{oder}^m$, mit $m \geq n$.
- Nach dem Pumping-Lemma gibt es eine Zahl $k \geq 1$, so dass $\text{entweder}^{n+i\cdot k}\text{oder}^m \in D'$, für alle $i \in \mathbb{N}$.

Ist Deutsch regulär?

- Es gibt also Sätze in D', in denen die Zahl der *entweder*s die Zahl der *oder*s übersteigt.
- Also gibt es auch Sätze in D, in denen die Zahl der *entweder*s die Zahl der *oder*s übersteigt.
- Das steht im Widerspruch zu der Beobachtung, dass zu jedem *entweder* ein anderes *oder* gehört.
- Also ist D keine reguläre Sprache.

Derartige Klammer-Konstruktionen wie das deutsche *entweder ... oder ...* gibt es vermutlich in allen natürlichen Sprachen.[1] Daher sind Typ-3-Grammatiken unzulänglich, um natürliche Sprachen vollständig zu beschreiben.

[1] Über die südamerikanische Sprache Pirahã wird behauptet, dass es dort keine derartigen Konstruktionen gebe, was aber umstritten ist; siehe http://de.wikipedia.org/wiki/Piraha.