Das Pumping-Lemma für reguläre Sprachen

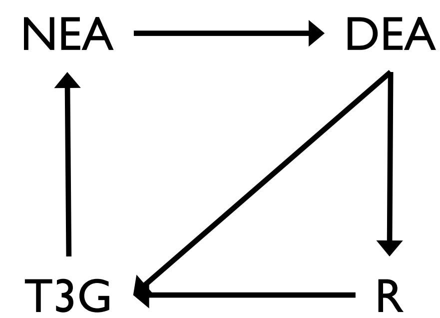
Chomsky-Hierarchie

Typ	Name	Erlaubte Produktionen	Akzeptierende Maschine	Beispiel
3	Regulär	N → wM $w \in T^*$	Endlicher Automat	a^n
2	Kontextfrei	N → w $w \in (N \cup T)^*$	Kellerautomat	$a^n b^n$
1	Kontext-sensitiv	uNv → uwv $u, v \in (N \cup T)^*$ $w \in (N \cup T)^+$ S → eps	Linear gebundener Automat	$a^n b^n c^n$
0	Rekursiv aufzählbar	u → v $u \in V^* N V^*, v \in V^*$ $V = (N \cup T)$	Turing Maschine	Halteproblem

Skript Worsch: Seite 80-83

Eine Sprachklasse

Erkenntnis:



- Typ-3-Grammatiken, deterministische endliche Automaten, nichtdeterministische endliche Automaten und Reguläre Ausdrücke beschreiben die gleiche Klasse von Sprachen.

- Es gibt Verfahren, die Beschreibung einer Sprache von einer dieser Definitionsarten in eine andere zu überführen.

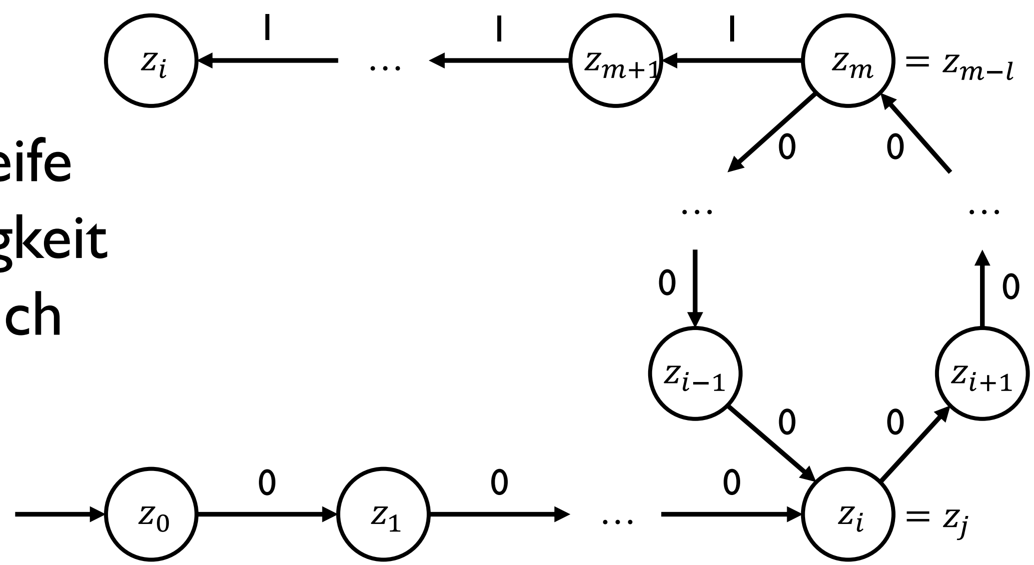
Abgrenzung dieser Sprachklasse

- Die Zugehörigkeit einer Sprache zu dieser Sprachklasse kann durch Angabe einer Typ-3-Grammatik, eines deterministischen endlichen Automaten, eines nichtdeterministischen endlichen Automaten oder eines Regulären Ausdrucks bewiesen werden.

- Wie beweist man die Nicht-Zugehörigkeit?

Begrenztheit des endlichen Automaten

- Ist ein Wort w lang genug (z.B. $|w| > |Z|$), müssen einige Zustände des endlichen Automaten mehrfach durchlaufen werden.
- Der Automat weiß nicht, wie häufig diese Zustände durchlaufen wurden.



$\Rightarrow$ Wörter, die diese Schleife in einer anderen Häufigkeit durchlaufen, müssen auch zur Sprache gehören!

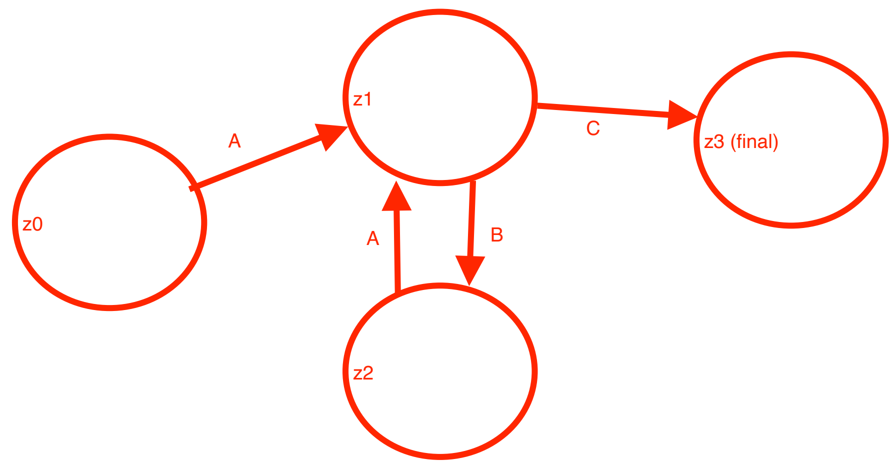
Begrenztheit erkennbarer Wörter

- Ist ein Wort w lang genug (z.B. $|w| > |Z|$), müssen einige Zustände des endlichen Automaten mehrfach durchlaufen werden.
- Das Wort w lässt sich in 3 Teile zerlegen $w = xyz$
  - x Präfix vor der Schleife
  - y Schleifeninhalt der sich beliebig oft wiederholen kann
  - z Postfix nach der Schleife
- Da die Schleife beliebig oft wiederholt werden kann sind auch die Wörter xy^kz Teil der Sprache

Formal ausgedrückt: Pumping-Lemma für reguläre Sprachen

- Sei L eine reguläre Sprache über einem Alphabet A.

- Dann gibt es eine Konstante n derart, so dass für alle Wörter $w \in L$ mit $|w| \geq n$ gilt:
- Es gibt Wörter $x, y, z \in A^*$ mit $w = xyz$, für die gilt:
  - $y \neq \varepsilon$
  - $|xy| \leq n$
  - $\forall k \in \mathbb{N}_0 \colon xy^kz \in L$



Beweis

- Sei L eine reguläre Sprache über einem Alphabet A.

- Dann gibt es einen deterministischen endlichen Automaten $N = (Z, z_0, A, f, F)$, für den gilt $L(N) = L$.

- Wähle $n = |Z|$.
- Sei $w = a_1 a_2 a_3 \dots a_m \in L$ mit $|w| = m \geq n$.
- Seien die z_i, durch w in N durchlaufenen Zustände:
  $f^{**}(z_o, w\ ) = f^{**}(z_o, a_1 a_2 a_3 \dots a_m) = z_0 z_1 \dots z_m$

Beweis

- Seien die z_i, durch w in N durchlaufenen Zustände:
$f^{**}(z_o, w\ ) = f^{**}(z_o, a_1 a_2 a_3 \dots a_m) = z_0 z_1 \dots z_m$

- Da $w = a_1 a_2 a_3 \dots a_m \in L$ gilt $z_m \in F$

- Da $m \geq n$, können die z_i nicht paarweise verschieden sein. Mindestens ein Zustand muss mindestens doppelt vorkommen (Schubfachprinzip)!

- Es gibt also zwei Indizes p und q so dass $z_p = z_q$.

Beweis

- Es gibt also zwei Indizes p und q mit $p \neq q$, so dass $z_p = z_q$.
  - Seien p und q das erste solche Paar.
  - Da es nur n verschiedene Zustände gibt muss gelten $0 \leq p < q \leq n$. (z_n ist der $(n+1)$-te Zustand!)

- Das Wort w kann somit folgend in $w = xyz$ zerlegt werden:
  - $x = a_1 a_2 \dots a_p$
  - $y = a_{p+1} a_{p+2} \dots a_q$
  - $z = a_{q+1} a_{q+2} \dots a_m$

- Zu Zeigen:
  - $y \neq \varepsilon$
  - $|xy| \leq n$
  - $\forall k \in \mathbb{N}_0 \colon xy^k z \in L$

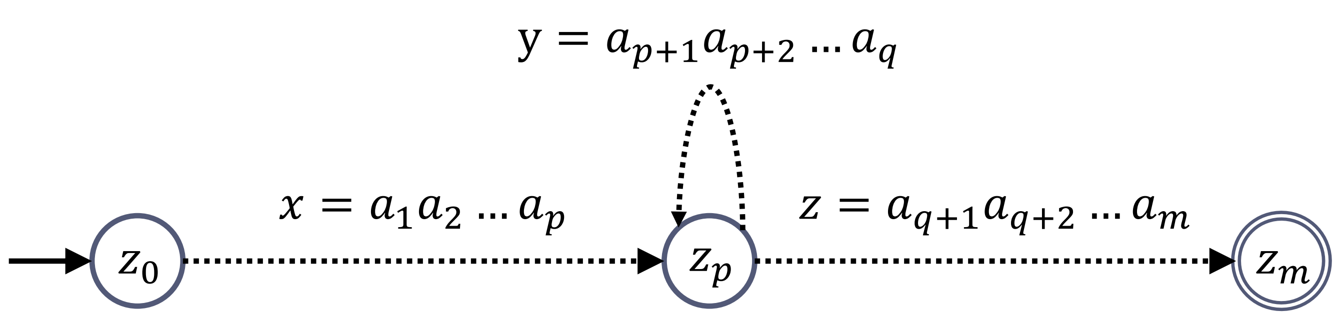
Beweis

- Das Wort w kann somit folgend in $w = xyz$ zerlegt werden:
  - $x = a_1 a_2 \dots a_p$
  - $y = a_{p+1} a_{p+2} \dots a_q$
  - $z = a_{q+1} a_{q+2} \dots a_m$
  - $0 \leq p < q \leq n$

- Zu Zeigen:
  - $y \neq \varepsilon$:
    Gegeben, da $p \neq q$
  - $|xy| \leq n$:
    Gegeben, da $|xy| = a_1 \dots a_p a_{p+1} \dots a_q$ und $q \leq n$
  - $\forall k \in \mathbb{N}_0 : xy^k z \in L$

Beweis ($\forall k \in \mathbb{N}_0 \colon xy^k z \in L$)

- Es gilt dann:
  - $f^*(z_0, x) = f^*(z_0, a_1 a_2 \dots a_p) = z_p$
  - $f^*(z_p, y) = f^*(z_p, a_{p+1} a_{p+2} \dots a_q) = z_p$
  - $f^*(z_p, z) = f^*(z_p, a_{q+1} a_{q+2} \dots a_m) = z_m$



Beweis $(\forall k \in \mathbb{N}_0 : xy^k z \in L)$

- Betrachte nun die Wörter $xy^k z$ mit $k \in \mathbb{N}_0$:
  - $f^*(z_0, x) = z_p$
  - $f^*(z_p, y^k) = z_p$
  - $f^*(z_p, z) = z_m$
  - $f^*(z_0, xy^k z) = z_m$
- Also gilt $\forall k \in \mathbb{N}_0 : xy^k z \in L$.

Beweis der Nicht-Zugehörigkeit

- Pumping-Lemma:
  - Wenn eine Sprache regulär ist und
  - ein hinreichend langes Wort in dieser Sprache ist,
  - dann kann ich das Wort so zerlegen dass auch weitere aus der Zerlegung konstruierte Wörter in dieser Sprache sind.

- Schön!
- Aber wie beweist man damit,
  dass eine Sprache nicht regulär ist?

Idee: Widerspruchsbeweis

- Pumping-Lemma:
  - Wenn eine Sprache regulär ist und
  - ein hinreichend langes Wort in dieser Sprache ist,
  - dann kann ich das Wort so zerlegen dass auch weitere aus der Zerlegung konstruierte (aufgepumpte) Wörter in dieser Sprache sind.

- Ansatz:
  - Wenn ein hinreichend langes Wort in einer Sprache ist und
  - Zu jeder Zerlegung des Worts ein aufgepumptes Wort existiert das nicht Teil der Sprache ist
  - Dann kann die Sprache nicht regulär sein.

Wiederholung: Pumping-Lemma für reguläre Sprachen

- Sei L eine reguläre Sprache über einem Alphabet A.

- Dann gibt es eine Konstante n derart, so dass für alle Wörter $w \in L$ mit $|w| \geq n$ gilt:
- Es gibt Wörter $x, y, z \in A^*$ mit $w = xyz$ für die gilt:

  - $y \neq \varepsilon$
  - $|xy| \leq n$
  - $\forall k \in \mathbb{N}_0 : xy^k z \in L$

Beispiel

Sei $L = \{0^t 1^t | t \in \mathbb{N}_0\}$

Zu zeigen:

- Egal wie n gewählt wird,
- gibt es ein Wort $w \in L$ mit $|w| \geq n$, für das es keine Zerlegung $w = xyz$ gibt, so dass gilt:
  - $y \neq \varepsilon$
  - $|xy| \leq n$
  - $\forall k \in \mathbb{N}_0: xy^k z \in L$

Beispiel - Lösungsansatz

Sei $L = \{0^t 1^t | t \in \mathbb{N}_0\}$

- Wähle w so, dass y nur Nullen enthält
- Dann verletzt eine Wiederholung von y die Bedingung dass es gleich viele 0 und 1 im Wort gibt

Beispiel – Wort das Pumping Lemma verletzt

- Sei $L = \{0^t 1^t | t \in \mathbb{N}_0\}$

- Sei n $\in \mathbb{N}$ beliebig.
- Wähle $w \in L$ mit $w = 0^n 1^n$.
- Da $|xy| \leq n$ besteht xy aus lauter Nullen
- Daher besteht y aus lauter Nullen
- Jede Wiederholung von y erhöht die Zahl der Nullen um $|y|$ ohne die Zahl der Einsen zu erhöhen
- Die geforderte Zerlegung $w = xyz$ ist nicht möglich
- Die Sprache ist nicht regulär

Vorgehensweise?

Pumping-Lemma

- Sei L eine reguläre Sprache über einem Alphabet A.

- Dann gibt es eine Konstante n derart, so dass
- für alle Wörter $w \in L$ mit $|w| \geq n$ gilt:
- Es gibt Wörter $x, y, z \in A^*$ mit $w = xyz$ für die gilt:
  - $y \neq \varepsilon$
  - $|xy| \leq n$
  - $\forall k \in \mathbb{N}_0: xy^k z \in L$

Widerspruchsbeweis

- Wir nehmen die gegebene Sprache.

- Wir zeigen für beliebiges n:
- Es gibt ein Wort $w \in L$ mit $|w| \geq n$, für das
- es keine Zerlegung $w = xyz$ gibt, bei der gleichzeitig gilt:
  - $y \neq \varepsilon$
  - $|xy| \leq n$
  - $\forall k \in \mathbb{N}_0: xy^k z \in L$

Vorgehensweise?

Pumping-Lemma

- Es gibt Wörter $x, y, z \in A^*$ mit $w = xyz$ für die gilt:
  - $y \neq \varepsilon$
  - $|xy| \leq n$
  - $\forall k \in \mathbb{N}_0 : xy^k z \in L$

Widerspruchsbeweis

- Für alle Wörter $x, y, z \in A^*$ mit $w = xyz$ für die gilt:
  - $y \neq \varepsilon$ und
  - $|xy| \leq n$,
  - $\exists\ k \in \mathbb{N}_0$ mit $xy^k z \notin L$

Beispiel

Widerspruchsbeweis

- Wir nehmen die gegebene Sprache.
- Wir zeigen für beliebiges n:
- Es gibt ein Wort, für das

Am Beispiel

- Sei $L_{prim} = \{1^t | \; t \; ist \; Primzahl\}$
- Sei n ∈ ℕ beliebig.
- Wir wählen eine Primzahl p Wir wählen also $w = 1^p$.

Beispiel

Widerspruchsbeweis

- Für alle Wörter $x, y, z \in A^*$ mit $w = xyz$ für die gilt
  - $y \neq \varepsilon$ und
  - $|xy| \leq n$,

- gibt es ein $k \in \mathbb{N}_0$ mit $xy^kz \notin L$

Am Beispiel

- Seien $x, y, z \in \{1\}^*$beliebig, so dass $w = 1^p = xyz$ und
  - $y \neq \varepsilon$ und
  - $|xy| \leq n$,

- Wir suchen k so dass $|xy^kz|$ keine Primzahl sein kann

Beispiel

Widerspruchsbeweis

- Für alle Wörter $x, y, z \in A^*$ mit $w = xyz$ für die gilt
  - $y \neq \varepsilon$ und
  - $|xy| \leq n$,

- gibt es ein $k \in \mathbb{N}_0$ mit $xy^k z \notin L$

Am Beispiel

- Seien $x, y, z \in \{1\}^*$ beliebig, so dass $w = 1^p = xyz$ und
  - $y \neq \varepsilon$ und
  - $|xy| \leq n$,

- Sei $|y| = m$.
  Wähle $k = p - m$.
  Dann gilt:
  $|xz| = p - m$
  $|xy^k z| = |xy^{p-m} z|$

Beispiel

Widerspruchsbeweis

- gibt es ein $k \in \mathbb{N}_0$ mit $xy^kz \notin L$

Am Beispiel

- $|xz| = p - m$

  $|xy^kz| = |xy^{p-m}z| =$
  $|xzy^{p-m}| = |xz| + |y^{p-m}| =$
  $p - m + m(p - m) =$
  $(p - m)(1 + m)$

Beispiel

Widerspruchsbeweis

- gibt es ein $k \in \mathbb{N}_0$ mit $xy^kz \notin L$

Am Beispiel

- $xy^kz = 1^{(p-m)(1+m)}$

  $(p - m)(1 + m)$ ist keine Primzahl!

  Es gilt somit $xy^kz \notin L$.

Beispiel – Hinweis für die ganz genauen

Wir wählen $p \geq n + 2$ damit $(p - m) \geq 2$

Zusammenfassung

- Verschiedene Möglichkeiten Typ-3-Sprachen zu definieren
  - Deterministische endliche Akzeptoren
  - Nichtdeterministische endliche Akzeptoren
  - Reguläre Ausdrücke
  - Rechts-(Links-)lineare Grammatiken
- Diese Methoden sind gleich mächtig
  - Wir können algorithmisch von jeder Methode in jede andere wechseln.
  - Jeder Methode „eignet“ sich für verschiedene Sprachen bzw. Anwendungsfälle.
- Mit Hilfe des Pumping-Lemmas kann in den meisten Fällen bewiesen werden, dass eine Sprache keine Typ-3-Sprache ist.

Das Pumping-Lemma für kontextfreie Sprachen