

Theoretische Grundlagen der Informatik

Tutorium 4

Institut für Kryptographie und Sicherheit



1. Produktionen auf Terminale und Nicht-Terminale sortieren
2. Produktionen auf mehr als zwei Nicht-Terminale ersetzen
3. Produktionen auf λ ersetzen
4. Produktionen auf ein Nicht-Terminal ersetzen
5. Neues Startsymbol einführen falls $S \rightarrow \lambda$ existiert
6. ???
7. PROFIT

Gegeben sei die folgende Grammatik: $\mathcal{G} = (\mathcal{T}, \mathcal{V}, S, \mathcal{P})$ mit

$\mathcal{T} := \{a, b, c, d\}$, $\mathcal{V} := \{S, A, D, M\}$,

$\mathcal{P} := \{S \rightarrow AMD \mid M, A \rightarrow AA \mid a, D \rightarrow DD \mid d, M \rightarrow bMc \mid \lambda\}$

1. Geben Sie die erzeugte Sprache an!
2. Wandeln Sie die gegebene kontextfreie Grammatik \mathcal{G} in eine äquivalente kontextfreie Grammatik \mathcal{G}' in Chomsky-Normalform um, indem sie jeden Schritt durch eine neue Grammatik beschreiben!
3. Zeigen oder widerlegen Sie mit Hilfe des CYK-Algorithmus, ob die folgenden Wörter in der Sprache \mathcal{L} liegen, die durch die Grammatik \mathcal{G} erzeugt wird!

3.1 *aabbccdd*

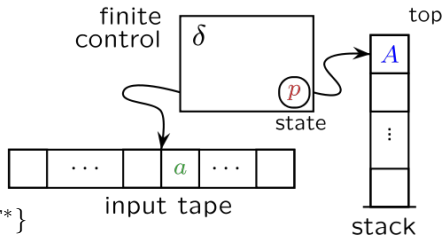
3.2 *abbcc*

3.3 *abcdd*

Definition Kellerautomaten

Ein (nichtdeterministischer) **Kellerautomat** (NPDA bzw PDA, Pushdown Automaton) besteht aus $(Q, \Sigma, \Gamma, q_0, Z_0, \delta, F)$, wobei

- Q endliche Zustandsmenge
- Σ endliches Eingabealphabet
- Γ endliches Stack-Alphabet
- $q_0 \in Q$ Anfangszustand
- $Z_0 \in \Gamma$ Initialisierung des Stacks
- $\delta : Q \times (\Sigma \cup \{\lambda\}) \times \Gamma \rightarrow 2^{Q \times \Gamma^*}$
 - $\delta(q, a, Z) \subseteq \{(q, \gamma) : q \in Q, \gamma \in \Gamma^*\}$
 - $\delta(q, \lambda, Z) \subseteq \{(q, \gamma) : q \in Q, \gamma \in \Gamma^*\}$
- $F \subseteq Q$ Menge der akzeptierenden Endzustände, $F = \emptyset$ ist möglich.



- Akzeptieren nach Eingabeende, wenn
 - der Stack leer ist *oder*
 - der Automat in einen akzeptierenden Zustand kommt.
- Sind im Allgemeinen nichtdeterministisch
- Man kann Endzustände auch aus der Definition weglassen und alternativ verlangen, dass der Automat genau bei leerem Keller akzeptiert.
- Man kann sogar alle Zustände bis auf einen weglassen und alles in die Kellerbelegung kodieren

$$M = (Q, \Sigma, \Gamma, q_0, Z_0, \delta, F)$$

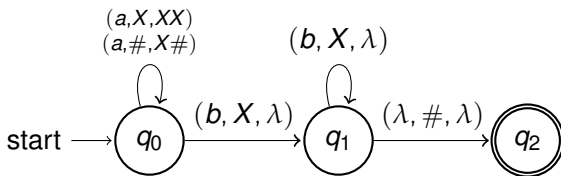
$$Q = \{q_0, q_1, q_2\}$$

$$\Sigma = \{a, b\}$$

$$\Gamma = \{\#, X\}$$

$$Z_0 = \#$$

$$F = \{q_2\}$$



- Welche Sprache akzeptiert dieser Automat?

Gegeben sei folgende Sprache für das Alphabet $\Sigma = \{a, b, c\}$:

$$\mathcal{L} = \{w_1 w_2 \in \Sigma^* \mid w_1 \in \{a, b\}^*, w_2 \in \{b, c\}^*, \\ \#_a w_1 + \#_b w_1 = \#_b w_2 + \#_c w_2\}$$

Hier gibt $\#_x w$ die Häufigkeit des Vorkommens eines Zeichens $x \in \Sigma$ in einem Wort $w \in \Sigma^*$ an.

1. Zeigen Sie, dass \mathcal{L} nicht regulär ist!
2. Geben Sie eine Chomsky-2-Grammatik an, die genau die Sprache \mathcal{L} erzeugt!
3. Geben Sie einen Kellerautomaten \mathcal{M} an, der genau die Sprache \mathcal{L} erkennt! Zeichnen Sie den Zustandsübergangsgraphen für \mathcal{M} !

Lemma

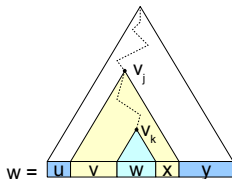
Für jede kontextfreie Sprache L gibt es eine Konstante $n \in \mathbb{N}$, so dass sich jedes Wort $z \in L$ mit $|z| \geq n$ so als

$$z = uvwxy$$

schreiben lässt, dass

- $|vx| \geq 1$,
- $|vwx| \leq n$ und
- für alle $i \geq 0$ das Wort $uv^iwx^iy \in L$ ist.

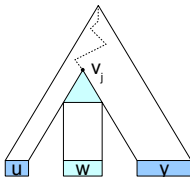
- Jeder Knoten im Ableitungsbaum (wie wir ihn in CYK sehen) steht für ein Nichtterminalsymbol
- Ab einer gewissen Höhe des Baumes (bzw. Länge des Wortes) muss ein Nichtterminal im Baum mehrmals in einer Reihe vorkommen
- Man kann also aus einem Nichtterminalsymbol dasselbe Symbol wieder ableiten
- Da das Wort durch jede Ableitung (außer zu Terminalsymbolen) länger wird, gibt es eine „Schleife“ beim Ableiten
- Diese Schleife kann man also „pumpen“, also beliebig oft (oder auch gar nicht) durchlaufen



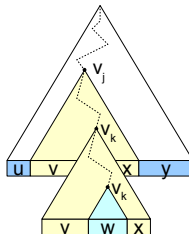
} h

Gegeben: Wort $z \in L$
mit $|x| \geq n$

Ableitungsbaum T für z
mit Höhe $h \geq N$



Erzeugen von uv^0wx^0y



Erzeugen von uv^2wx^2y

Pumping Lemma Formalia (kontextfrei)

Behauptung: L ist nicht kontextfrei.

Beweis:

Nehme an L sei kontextfrei.

Sei n beliebig aber fest.

Wähle $z = \text{_____} \in L$ mit $|z| \geq n$

Beh.: $\forall u, v, w, x, y : uvwxy = z$ mit $|vx| \geq 1$ und $|vwx| \leq n$, $\exists i \in \mathbb{N}$,
so dass $uv^iwx^iy \notin L$.

Bew.: _____

Widerspruch zum Pumping Lemma $\Rightarrow L$ ist nicht kontextfrei.

Zeige, dass die Sprache



$$L = \{\omega\omega \mid \omega \in \{0, 1\}^*\}$$

nicht kontextfrei ist.

Aufgabe 1

1. Geben Sie für die Sprache $\mathcal{L} = \{a^n b^n c^n \mid n \in \mathbb{N}\}$ eine Grammatik des höchstmöglichen Chomsky-Typs an!
2. Zeigen Sie, dass die Sprache $\mathcal{L}' = \{a^{2^n} \mid n \in \mathbb{N}\}$ nicht kontextfrei ist!

Bis zum nächsten Mal!

<p>□□□□□□□□□□□□□□</p> <p>UNCOMMON (NON-GIBBERISH) BASE WORD</p> <p>ORDER UNKNOWN</p> <p>Trøub4dor &3</p> <p>CAPS? COMMON SUBSTITUTIONS NUMERICAL PUNCTUATION</p> <p>(YOU CAN ADD A FEW MORE BITS TO FREQUENCY FOR THE FACT THAT THIS IS ONLY ONE OF A FEW COMMON FORMATS)</p>	<p>~28 BITS OF ENTROPY</p> <p>□□□□□□□□ □□□□□□□□ □□□□ □□□□ □□□□</p> <p>$2^{28} = 3 \text{ DAYS AT } 1000 \text{ GUESSES/SEC}$</p> <p>(PLAUSIBLE ATTACK ON A WORK REMOTE WEB SERVICE YES, CRACKING A STOLEN MP3 IS FASTER, BUT IT'S NOT WHAT THE AVERAGE USER SHOULD WORRY ABOUT.)</p> <p>DIFFICULTY TO GUESS: EASY</p>	<p>WAS IT TROMBONE? NO, TROUBADOR. AND ONE OF THE O'S WAS A ZERO?</p> <p>AND THERE WAS SOME SYMBOL...</p>  <p>DIFFICULTY TO REMEMBER: HARD</p>
<p>correct horse battery staple</p> <p>□□□□□□ □□□□□□ □□□□□□ □□□□□□</p> <p>FOUR RANDOM COMMON WORDS</p>	<p>~44 BITS OF ENTROPY</p> <p>□□□□□□□□□□ □□□□□□□□□□ □□□□□□□□□□ □□□□□□□□□□</p> <p>$2^{44} = 530 \text{ YEARS AT } 1000 \text{ GUESSES/SEC}$</p> <p>DIFFICULTY TO GUESS: HARD</p>	<p>THAT'S A BATTERY STAPLE.</p> <p>CORRECT!</p>  <p>DIFFICULTY TO REMEMBER: YOU'VE ALREADY MEMORIZED IT</p>

THROUGH 20 YEARS OF EFFORT, WE'VE SUCCESSFULLY TRAINED EVERYONE TO USE PASSWORDS THAT ARE HARD FOR HUMANS TO REMEMBER, BUT EASY FOR COMPUTERS TO GUESS.



Dieses Werk ist unter einem "Creative Commons Namensnennung-Weitergabe unter gleichen Bedingungen 3.0 Deutschland"-Lizenzvertrag lizenziert. Um eine Kopie der Lizenz zu erhalten, gehen Sie bitte zu <http://creativecommons.org/licenses/by-sa/3.0/de/> oder schreiben Sie an Creative Commons, 171 Second Street, Suite 300, San Francisco, California 94105, USA.

Davon ausgenommen sind das Titelbild, welches aus der März-April 2002 Ausgabe von American Scientist erschienen ist und ohne Erlaubnis verwendet wird, sowie das KIT Beamer Theme. Hierfür gelten die Bestimmungen der jeweiligen Urheber.