Prof. J. Křetínský
M. Weininger, T. Meggendorfer

Technische Universität München

KEINE ABGABE

Einführung in die theoretische Informatik

Sommersemester 2018 – Übungsblatt 7

Selbstständige Vorbereitung

Bereiten Sie sich auf die Tutorgruppen selbstständig vor, indem Sie die Aufgaben 7.1 bis 7.4 ansehen. Insbesondere Aufgabe 7.4 sollten Sie lösen, da Aufgabe 7.7 darauf aufbaut.

AUFGABE 7.1. (*Wichtige Begriffe*) Stufe A

Überprüfen Sie, dass Sie die Folgenden Begriffe korrekt definieren können.

- Chomsky-Normalform
- Abschlusseigenschaften kontextfreier Sprachen
- CYK-Algorithmus
- Pumping-Lemma für kontextfreie Sprachen

AUFGABE 7.2. (*Chomsky-Normalform üben*) Stufe C

Rufen Sie die Website `http://grammar.epfl.ch/` auf und üben Sie das Überführen von Grammatiken in Chomsky-Normalform.

AUFGABE 7.3. (*Komplemente von CFL*) Stufe C

Aus der Vorlesung wissen Sie bereits, dass kontextfreie Sprachen *nicht* über Schnitt und Komplement abgeschlossen sind. In dieser Aufgabe geht es darum, zu sehen, dass es dennoch kontextfreie, nicht-reguläre Sprachen gibt, deren Komplement ebenfalls kontextfrei und nicht regulär ist. Sei $\mathsf{L} = \{\mathsf{a}^n\mathsf{b}^n\mathsf{c}^m \mid n, m \in \mathbb{N}_0\}$ über dem Alphabet $\Sigma = \{\mathsf{a}, \mathsf{b}, \mathsf{c}\}$ eine nicht-reguläre, kontextfreie Sprache.

(a) Geben Sie eine kontextfreie Grammatik G_1 für L an.
(b) Geben Sie eine kontextfreie Grammatik G_2 für $\overline{\mathsf{L}}$ an.

Hinweis: Sie dürfen auch andere ε-Produktionen als $\mathsf{S} \to \varepsilon$ verwenden. Warum?

AUFGABE 7.4. (*CYK-Algorithmus*) Stufe C

Wir betrachten die Grammatik $\mathsf{G} = (\{\mathsf{S}, T, U, A, B, C\}, \{\mathsf{a}, \mathsf{b}, \mathsf{c}\}, \mathsf{P}, \mathsf{S})$ in CNF mit den folgenden Produktionen P:

$$\begin{aligned} \mathsf{S} &\to T\mathsf{S} \mid CT \mid \mathsf{a} & \qquad A &\to \mathsf{a} \\ T &\to AU \mid TT \mid \mathsf{c} & B &\to \mathsf{b} \\ U &\to \mathsf{S}B \mid AB & C &\to \mathsf{c} \end{aligned}$$

(a) Bestimmen Sie mit dem CYK-Algorithmus, ob $\mathsf{ccaab} \in \mathsf{L}(\mathsf{G})$ und $\mathsf{aabcc} \in \mathsf{L}(\mathsf{G})$. Geben Sie dabei auch die berechneten Tabellen an.
(b) Beschreiben Sie eine Erweiterung des CYK-Algorithmus, mit welcher für ein gegebenes $w \in \mathsf{L}(\mathsf{G})$ alle Ableitungsbäume bzgl. G berechnet werden können, und wenden Sie dieses Verfahren auf die Wörter aus (a) an.

AUFGABE 7.5. (*Pumping-Lemma für Kontextfreie Sprachen*) Stufe C

Wir betrachten Pfeil-Sprachen über dem Alphabet $\Sigma = \{\uparrow, \downarrow, \leftarrow, \rightarrow\}$. Wir interpretieren dabei ein Wort $w \in \Sigma^*$ als einen Pfad in einem 2D-Gitter.

(a) Geben Sie eine formale Definition für die folgenden beiden Sprachen an:[1]
   (i) Pfade, die in den Ursprung zurückkehren.
   (ii) Pfade, die „in großer Kurve umkehren" — *beliebig weit* nach rechts fahren, dann *noch weiter* entweder nach oben oder unten gehen und letztlich wieder umkehren und *noch weiter* nach links fahren. Diese Pfade enden dann links vom Ursprung.
(b) Zeigen Sie mit Hilfe des Pumping-Lemma für kontextfreie Sprachen, dass diese Sprache nicht kontextfrei sind

AUFGABE 7.6. (*Abschlusseigenschaften kontextfreier Sprachen*) Stufe D

Beweisen Sie die folgenden beiden Aussagen:

(a) Der Präfixabschluss $\mathsf{L}_{pre} \coloneqq \{u \mid \exists v.\ uv \in \mathsf{L}\}$ einer kontextfreien Sprache $\mathsf{L} \subseteq \Sigma^*$ über einem Alphabet Σ ist wieder kontextfrei. Geben Sie hierzu ein Übersetzung an, die die Grammatik G für L in eine Grammatik G' umwandelt, sodass $\mathsf{L}_{pre} = \mathsf{L}(G')$ übersetzt.
(b) Kontextfreie Sprachen sind unter Schnitt mit regulären Sprachen abgeschlossen.

Stufe C/D

AUFGABE 7.7. (*CYK-Algorithmus — Teil 2*)
Sei G durch folgende Produktionen gegeben:

$$\mathsf{S} \to AB \mid CD \mid AT \mid CU \mid \mathsf{SS} \qquad T \to \mathsf{S}B \qquad U \to \mathsf{S}D \qquad A \to (\qquad B \to) \qquad C \to \{ \qquad D \to \}$$

(a) Wenden Sie den CYK-Algorithmus auf folgende Wörter an:

$$w_1 = ()\{()\} \qquad w_2 = ()\{(\qquad w_3 = ()\},$$

indem Sie die Tabellen, die Sie mithilfe des CYK-Algorithmus berechnen, angeben.

(b) Diskutieren Sie, wie man beim CYK-Algorithmus das mehrmalige Berechnen der gleichen Tabelle verhindern kann. Bestimmen Sie dazu, in welcher Beziehung die Worte w_1 bis w_3 zueinander stehen und warum es in Aufgabenteil (a) reicht, eine einzige Tabelle zu berechnen.

(c) Wir möchten den CYK-Algorithmus so abändern, dass er Eingabefehler (z.B. eine öffnende Klammer ohne passende schließende Klammer) erkennt und die Position des Fehlers bestimmt. *Geben Sie hierzu eine neue Variante des CYK-Algorithmus so an, dass er für gegebenes $w \in \Sigma^*$ den längsten Präfix u bestimmt, so dass sich u zu einem Wort in $\mathsf{L}(\mathsf{G})$ ergänzen lässt.*
Beispiel:

$$\mathsf{G} \colon \mathsf{S} \to AB \mid \mathsf{SS} \qquad A \to \mathsf{a} \qquad B \to \mathsf{b}$$

Dann gilt $w = \mathsf{ababb} \notin \mathsf{L}(\mathsf{G})$, aber $\mathsf{ab}, \mathsf{abab} \in \mathsf{L}(\mathsf{G})$; weiterhin gibt es kein $v \in \Sigma^*$, so dass $wv \in \mathsf{L}(\mathsf{G})$ gilt, womit abab das gesuchte Präfix ist. Im Fall $w = \mathsf{aba}$ wäre aba das gesuchte Präfix, da $w\mathsf{b} \in \mathsf{L}(\mathsf{G})$ gilt.

(d) Wenden Sie den Algorithmus aus (c) auf die Wörter aus (a) an, um das maximalen Präfix zu bestimmen, der sich noch zu einem Wort in $\mathsf{L}(\mathsf{G})$ ergänzen lässt.

AUFGABE 7.8. Stufe E

Sei Σ ein Alphabet und $\mathsf{L} \subseteq \Sigma^*$. L ist *präfix-abgeschlossen*, wenn für alle $w \in \mathsf{L}$ und $v \in \Sigma^*$ mit $v \trianglelefteq w$ (also v ist ein Präfix von w) gilt: $v \in \mathsf{L}$.
Beweisen Sie, dass jede unendliche, präfix-abgeschlossene, kontextfreie Sprache eine unendliche, reguläre Teilmenge enthält.

AUFGABE 7.9. (*Ogdens Lemma*) Stufe E

Sei L eine CFL. Dann existiert ein $p \in \mathbb{N}_0$, so dass für jedes $z \in \mathsf{L}$ mit $|z| \geq p$ und jede *Markierung* von mindestens p Zeichen in z gilt: Es gibt eine Zerlegung $z = uvwxy$ mit

- vx enthält mindestens ein markiertes Zeichen.
- vwx enthält höchstens p markierte Zeichen.
- $uv^iwx^iy \in \mathsf{L}$ für jedes $i \in \mathbb{N}_0$.

Eine geeignete Markierung kann z.B. Unterstreichung $\underline{\mathsf{a}}$, Apostroph a' oder Einfärbung a sein. Markiert man stets alle Zeichen, so erhält man das ursprüngliche Pumping-Lemma für CFL.

(a) Sei $\mathsf{L} = \{a^ib^jc^kd^l \mid i = 0 \vee j = k = l\}$.
Zeigen Sie mittels Ogdens Lemma, dass L nicht kontextfrei ist.
(Warum reicht das Pumping-Lemma für CFL nicht aus, um zu zeigen, dass L nicht kontextfrei ist?)

(b) Beweisen Sie Ogdens Lemma.

[1] Wenn Ihnen nicht direkt eine formale Definition einfällt, gehen Sie wie wir auf Blatt 6 vor, d.h. machen Sie Beispiele von Wörtern, die in der Sprache enthalten bzw. nicht enthalten sind.