Theoretische Informatik III (T3INF2002)

Formale Sprachen und Automaten | Einführung Compilerbau

Übungseinheit im Wintersemester 2022/23

Formale Sprachen und Automaten

- Übungen zu formalen Sprachen
- Übungen zu Pumping-Lemma

Entscheidungsprobleme

Wortproblem

- Bezeichnet das Problem, zu entscheiden, ob ein gegebenes Wort zur Sprache gehört oder nicht
- Wortproblem ist entscheidbar, wenn es einen Algorithmus gibt, der in endlicher Zeit herausfindet,
  ob $w \in L$ ist oder nicht
- Für die Sprachklassen nach Chomsky gilt:
  - Wortproblem für Typ-0-Sprachen ist rekursiv aufzählbar und nicht entscheidbar
  - Wortproblem für Typ-1-Sprachen ist entscheidbar (Zeitbedarf höchstens exponentiell)
  - Wortproblem für Typ-2-Sprachen ist durch den Earley-Algorithmus oder den Cocke-Younger-Kasami-Algorithmus entscheidbar(Zeitbedarf kubisch)
  - Wortproblem für Typ-3-Sprachen ist durch deterministische endliche Automaten lösbar (Zeitkomplexität ist linear)

Leerheitsproblem

- Bezeichnet das Problem, zu entscheiden, ob eine formale Sprache L leer ist ($L = \emptyset$) oder nicht
- Problem: Ermittlung von Wörtern die den Regeln der Grammatik genügen oder nicht
- Entscheidbarkeit des Leerheitsproblems hängt von der Komplexität der Grammatik der formalen Sprache ab
- Es gilt: für Grammatiken vom Typ 2 oder höher der Chomsky-Hierarchie ist das Leerheitsproblem entscheidbar, für Grammatiken bis Typ 1 jedoch nicht

Endlichkeitsproblem

- Bezeichnet das Problem, zu entscheiden, ob die Sprache endlich ist
- Eine formale Sprache wird als endlich bezeichnet, wenn die Menge ihrer „Wörter“ endlich ist ($|L| < \infty$)
- Für reguläre und kontextfreie Sprachen ist das Endlichkeitsproblem entscheidbar, für Sprachen vom Typ-1 und Typ-0 der Chomsky-Hierarchie jedoch nicht

Äquivalenzproblem

- Bezeichnet das Problem, zu entscheiden, ob zwei formale Definitionen von den Sprachen L_1 und L_2 äquivalent sind ($L_1 = L_2$)
- Äquivalenzproblem ist für reguläre Grammatiken und deterministische kontextfreie Grammatiken entscheidbar -> für nicht-deterministische kontextfreie Grammatiken hingegen nicht

Entscheidungsprobleme regulärer Sprachen

Problem	Eingabe	Fragestellung	Entscheidbar?
Wortproblem	Sprache L, Wort $\omega \in \Sigma^*$	Ist $\omega \in L$?	Ja
Leerheitsproblem	Sprache L	Ist $L = \emptyset$?	Ja
Endlichkeitsproblem	Sprache L	Ist $\|L\| < \infty$?	Ja
Äquivalenzproblem	Sprachen L_1 und L_2	Ist $L_1 = L_2$?	Ja

Pumping-Lemma für reguläre Sprachen

Pumping-Lemma

- Wird verwendet, um einen Widerspruchsbeweis zu führen, der hilft zu entscheiden, ob es sich bei einer Sprache um eine reguläre Sprache handelt
- Es gibt eine Mindestlänge n, durch die sich Wörter $x \in L$, mit $|x| \geq n$, in x = uvw zerlegen lassen
- Die Zerlegung erfüllt die folgenden Eigenschaften:

    1. $|v| \geq 1$
    2. $|uv| \leq n$
    3. für alle $\mathrm{i} \geq 0$: $\mathrm{uv^iw} \in L$

- Widerspruchbeweis in 4 Schritten

Beispiel $L_{C2} := \{a^n b^n \mid n \in N^+\}$: Schritt 1

- Annahme L sei regulär.
- Dann gibt es nach dem Pumping-Lemma eine Mindestlänge n, sodass sich alle Wörter $x \in L$ mit mindestens der Länge n zerlegen lassen in x = uvw und dabei die drei Eigenschaften des Pumping-Lemmas erfüllen.

Beispiel L_{C2} :={$a^n b^n$ | n ∈ N^+}: Schritt 2

- Wähle ein Wort in der Sprache L ($x \in L$) -> x = $a^n b^n$
- Die Länge $|x| = 2n$ des gewählten Wortes x ist größer als die Mindestlänge $|x| \geq n$
- Wort ist demnach geeignet

Beispiel L_{C2} :=$\{a^n b^n \mid n \in N^+\}$: Schritt 3

— Aufteilung des Wortes x in u v w

Fall 1: Der mittlere Wortteil v besteht nur aus a-Symbolen: uvw = $a^k a^m b^n$
mit $k + m = n$

Fall 2: Der mittlere Wortteil v besteht nur aus b-Symbolen: uvw = $a^k b^m b^n$
mit $k + m = n$

Fall 3: Der mittlere Wortteil v besteht aus a- und b-Symbolen: uvw = $a^k a^m b^s b^r$
mit $k + m = n$ und $s + r = n$

Beispiel L_{C2} :={$a^n b^n$ | n ∈ N^+}: Schritt 4

- Zeigen, dass mindestens eine Eigenschaft des Pumping-Lemmas durch die Aufteilung verletzt wird
- Wahl von i = 2, um das Wort „aufzupumpen"

Fall 1: Der mittlere Wortteil v besteht nur aus a-Symbolen: $uv^2w = a^k a^{2m} b^n$
mit k + m = n

Offensichtlich ist k + 2m größer als n, denn k + m = n. Es gibt also mehr a- als b-Symbole. Damit ist die dritte Eigenschaft verletzt, denn das Wort $a^k a^{2m} b^n$ gehört nicht zur Sprache L.

Beispiel L_{C2} :={$a^n b^n$ | n ∈ N^+}: Schritt 4

- Zeigen, dass mindestens eine Eigenschaft des Pumping-Lemmas durch die Aufteilung verletzt wird
- Wahl von i = 2, um das Wort „aufzupumpen“

Fall 2: Der mittlere Wortteil v besteht nur aus b-Symbolen: $uv^2w = a^k b^{2m} b^n$
mit $k + m = n$

Offensichtlich ist 2m + k größer als n, denn m + k = n. Es gibt also mehr b- als a-Symbole. Damit ist die dritte Eigenschaft verletzt, denn das Wort $a^k b^{2m} b^n$ gehört nicht zur Sprache L.

Beispiel L_{C2} :={$a^n b^n$ | $n \in N^+$}: Schritt 4

- Zeigen, dass mindestens eine Eigenschaft des Pumping-Lemmas durch die Aufteilung verletzt wird
- Wahl von i = 2, um das Wort „aufzupumpen"

Fall 3: Der mittlere Wortteil v besteht aus a- und b-Symbolen: $uv^2w = a^k(a^m b^s\, a^m b^s)b^r$
mit $k + m = n$ und $s + r = n$

Das aufgepumpte Wort ist nicht in der Sprache, weil nach dem b-Symbol wieder ein a-Symbol folgt. Damit ist die dritte Eigenschaft verletzt.

-> Damit ist gezeigt, dass keine der möglichen Aufteilungen uvw gleichzeitig alle drei Eigenschaften des Pumping-Lemmas erfüllt. -> Widerspruch zu der Annahme, dass L regulär ist. Demnach ist L nicht regulär.

Beispiel

Übungsaufgaben

Aufgabe 1

Gegeben sei die folgende Grammatik: G = (T , V, S, P) mit T:={a, b, c, d}, V:={S, A, D, M},
P:={S → AMD | M,A → AA | a, D → DD | d, M → bMc | ε}

Geben Sie die erzeugte Sprache L an!

Hinweis: überlegen Sie sich zunächst Wörter, die sich aus der Grammatik erzeugen lassen

Aufgabe 2

Gegeben sei die Sprache $L = \{w \in \{a, b\}^* \mid w \text{ enthält gleich viele a wie b}\}$

Zeigen Sie mit Hilfe des Pumping-Lemmas, dass L nicht regulär ist!

Aufgabe 3

Gegeben sei die Sprache $L = \{wcw^R \mid w \in \{a, b\}^*\}$[1]

Zeigen Sie mit Hilfe des Pumping-Lemmas, dass L nicht regulär ist!

Aufgabe 4

Gegeben ist die Sprache

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

für das Alphabet $\Sigma = \{a, b, c\}$

-> $\#_x w$ Häufigkeit des Vorkommens eines Zeichens $x \in \Sigma$ in einem Wort $w \in \Sigma^*$ an

1. Zeigen Sie, dass L nicht regulär ist.

2. Geben Sie eine Chomsky-2-Grammatik an, durch die die Sprache L erzeugt werden kann.