Übungsblatt 4
Sprachen, Grammatiken und reguläre Ausdrücke

Theoretische Informatik
Studiengang Angewandte Informatik
Wintersemester 2015/2016
Prof. Barbara Staehle, HTWG Konstanz

DIESE ÜBUNG WURDE AUF RICHTIGKEIT NICHT GEPRÜFT

Aufgabe 4.1

Pumping-Lemma

Teilaufgabe 4.1.1

Pumping-Lemma für reguläre Sprachen, credits = 2, label = pumpingReg

In der Vorlesung wurde mit Hilfe des Pumping-Lemmas bewiesen, dass die Sprache $L_{C_2} = \{a^n b^n | n \in N\}$ nicht regulär ist.

Können Sie diesen Beweis nutzen, um zu entscheiden, ob die Sprache

$$L_1 = \{\omega \in \{a, b\}^* || \omega|_a = |\omega|_b\} = \{ab, ba, \ldots, babbaaabab, \ldots\}$$

(also die Sprache, welche alle Wörter mit gleich vielen as und bs enthält) regulär ist?

Egal ob Sie diesen Beweis nutzen oder nicht - entscheiden Sie, ob L_1 regulär ist oder nicht und begründen Sie dies.

Lösung :

angenommen ist sie regulär, betrachten wir zunächst das in L_1 enthaltene Wort $w = b^n a^m$ mit $m = n$. wir zerlegen das obige Wort in :

Fall 1 :

$$u = b^i \quad v = b^j \quad w = b^k a \quad \text{Mit} \quad i + j + k = m$$

Das Aufpumpen von v führt zum Widerspruch $uv^2w = b^i(b^j)^2 b^k a = b^i b^{2j} b^k a$

denn i+2j+k $\neq$ m

Fall 2 :

$$u = b^i \quad v = b^j a \quad w = \epsilon \quad \text{Mit} \quad i + j = m$$

Das Aufpumpen von v führt zum Widerspruch $uv^2w = b^i(b^ja)^2 = b^ib^{2j}a$ denn i+2j $\neq$ m
Das heißt dass die Sprache nicht regulär ist.

Teilaufgabe 4.1.2

topic = Pumping-Lemma für kontextfreie Sprachen, credits = 2, label = pumpingKF] In der Vorlesung wurde mit Hilfe des Pumping-Lemmas bewiesen, dass die Sprache $L_{C_1} = \{a^nb^nc^n|n \in \mathbf{N}\}$ nicht kontextfrei ist.

Können Sie diesen Beweis nutzen, um zu entscheiden, ob die Sprache

$$L_2 = \{a^ib^jc^k|i \leq j \leq k, i, j, k \in \mathbf{N}\} = \{abbcccc, cba, \ldots, bcccbacabacb, \ldots\}$$

(also die Sprache, welche alle Wörter mit höchstens so vielen as wie bs und höchstens so vielen bs wie cs enthält) kontextfrei ist?

Egal ob Sie diesen Beweis nutzen oder nicht - entscheiden Sie, ob L_2 kontextfrei ist oder nicht und begründen Sie dies.

Lösung :

angenommen ist sie kontextfrei, betrachten wir zunächst das in L_2 enthaltene Wort $w = abbcccc$. wir zerlegen das obige Wort in :
Fall :

$$u = ab^l \quad v = b^m \quad w = b^n \quad x = b^pcc \quad y = cc \quad \text{Mit} \quad l + m + n + p = j$$

Das Aufpumpen von v führt zum Widerspruch $uv^0wx^0y = ab^lb^pcc$ denn l+p $\neq$ j
Ausßerdem ist in diesem Wort $|j| = |k|$, was die Bedingung verletzt.
Das heißt dass die Sprache nicht regulär ist.

Aufgabe 4.2

topic = Eine CNF für die Dyck-Sprache, credits = 3

Aus der Vorlesung ist Ihnen die folgende Grammatik zur Erzeugung der Dyck-Sprache D_2 bekannt:

$$S \to \varepsilon | SS | [S] | (S)$$

1. Modifizieren Sie die Grammatik so, dass sich das leere Wort nicht mehr ableiten lässt. **Lösung:**

$$S \to SS | [S] | (S)$$

2. Übersetzen Sie die modifizierte Grammatik in Chomsky-Normalform. **Lösung:**

   (a) Elimination

   $$S \to SS | S | S | [S] | (S) | [] | ()$$

   (b) Kettenregel

   $$S \to SS | [S] | (S) | [] | () | [S] | (S) | [] | () | [S] | (S) | [] | ()$$

   (c) Separation

   $$S \to SS | N_[S N_] | N_(S N_) | [] | ()$$

   $$N_[\to [$$

   $$N_] \to]$$

   $$N_(\to ($$

   $$N_) \to)$$

   (d) Elimination von mehrelementigen Nonterminalkette

   nix zu tun

3. Generieren Sie mit Hilfe der Grammatik in Chomsky-Normalform eine Ableitung und den dazugehörigen Ableitungsbaum für das Wort ([])[] **Lösung:** siehe Blatt

Aufgabe 4.3

topic=Reguläre Ausdrücke Um diese Aufgabe lösen zu können, verwenden Sie die in der Vorlesung verwendete Datei $thInf_2 - formaleSprachen_beispiel_addresslist.txt$ welche in Moodle zur Verfügung steht. Außerdem benötigen Sie eine Linux Unix Shell, Cygwin unter Windows oder eine andere Windows-Grep-Lösung, oder ein Online-Tool wie z.B. Grep Online.

Teilaufgabe 4.3.1

topic = Städte, credits = 3] Finden Sie alle Zeilen, welche die Städte der Tourismus-Büros enthält.

Beispiel: „Montgomery, AL 36103“

Teilaufgabe 4.3.2

topic = Staaten und Büros, credits = 4] Finden Sie alle Zeilen, welche den Staat des Tourismus-Büros sowie dessen Namen enthält.

Beispiel: „Alabama Alabama Bureau of Tourism & Travel“

Lösung:

Montgomery, AL 36103
Juneau, AK 99811-0801
Phoenix, AZ 85007
Little Rock, AR 72201
Sacramento, CA 95812
Denver, CO 80202
Hartford, CT 06103
Dover, DE 19903
Tallahassee, FL 32302
Honolulu, HI 96804
Boise, ID 83720-0093
Indianapolis, IN 46204
Baton Rouge, LA 70804-9291
Augusta, ME 04333-0059
Baltimore, MD 21202
Boston, MA 02116
Jefferson City, MO 65102

Lincoln, NE 68509-8907
Carson City, NV 89701
Concord, NH 03302
Santa Fe, NM 87501
Albany, NY 12220-0603
Bismark, ND 58505-0825
Pierre, SD 57501-5070
Richmond, VA 23219
Charleston, WV 25305
Madison, WI 53708-8690
Cheyenne, WY 82002

Aufgabe 4.4

[topic = Typ-1- und Typ-0-Sprachen, credits = 4]

Geben Sie (nicht in der Vorlesung erwähnte) Grammatiken G_1 und G_0 an, welche zu den Chomsky-Typen 1 bzw. 0 gehören.

Verwenden Sie das Terminalalphabet $\Sigma = \{a, b\}$ und begründen Sie (bei Bedarf auch anhand der Sprachen $L_1 = \mathcal{L}(G_1)$ bzw. $L_0 = \mathcal{L}(G_0)$), weshalb Ihre Grammatiken nicht auch von einem höheren Chomsky-Typ sind.

Lösung: $G_0 = (\{S, B\}, \{a, b\}, P, S)$ mit P=
$S \to \epsilon | aSb$
$aS \to aB$
$B \to bB$
$B \to b$

Die kann nicht vom Typ 1 sein, wegen der Regel $S \to aSb$ da $S \to \epsilon$ existiert kann S nicht auf der rechten Seite einer anderen Grammatik stehen.Also Typ 0

$G_1 = (\{S, B\}, \{a, b\}, P, S)$ mit P=
$S \to \epsilon$
$aS \to aB$
$B \to bB$
$B \to b$

Die ist vom Typ 1, nicht vom Typ 2 wegen der Grammatik $aS \to aB$