Proseminar Theoretische Informatik

Deterministisch kontextfreie Sprachen

Boris Dimitrov **Dozent:Wolfgang Mulzer**

1. Definition

Deterministisch kontextfreie Sprachen sind die Sprachen, die von einem deterministischen Kellerautomat akzeptiert werden.

Obwohl die Kellerautomaten nichtdeterministisch sein können ist die Teilklasse der deterministischen recht wichtig. Insbesondere Parser verhalten sich wie deterministische Kellerautomaten. Daher ist die Klasse der Sprachen, die von diesen Automaten akzeptiert wird, von Interesse, weil sie uns Aufschluss über die Arten von Konstrukten gibt, die sich für die Verwendung in Programmiersprachen eignen.

2. Deterministisch kontextfreie Kellerautomaten

Ein Kellerautomat ist deterministisch, wenn in keiner Situation mehrere Bewegungen möglich sind.

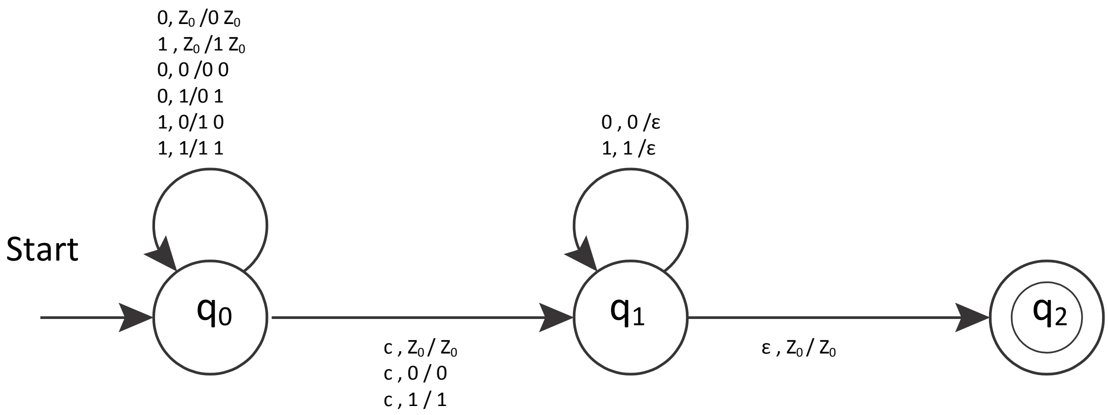
Wenn δ (q,a,x) mehr als ein Paar enthält, dann ist der Kellerautomat mit Sicherheit nichtdeterministisch, weil eines dieser Paare zur Bestimmung der nächsten Bewegung gewählt werden kann. Auch wenn δ (q,a,x) stets nur ein einzelnes Paar enthält, bestünde im Allgemeinen immer noch die Wahlmöglichkeit zwischen einem Eingabesymbol und einer ε-Bewegung.

Deswegen werden wir einen Kellerautomat als deterministisch bezeichnen, wenn folgende 2 Bedingungen erfüllt sind:

1. δ(q,a,x) besitzt für jedes q aus Q, a aus ∑ oder a = ε und x aus Γ höchstens ein Element.
2. Wenn δ(q,a,x) für ein a aus ∑ nicht leer ist, dann muss δ(q, ε, x) leer sein.

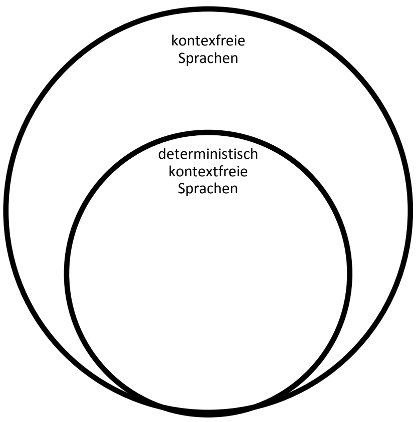
<u>Beispiel:</u>

Die Sprache L_{wcwr} ist eine Sprache, die von einem deterministischen Kellerautomat akzeptiert wird => ist eine deterministische kontextfreie Sprache.
Für jedes δ(q,a,x) des Automats (mit q aus Q, a aus ∑ und x aus Γ) ist die nächste Bewegung eindeutig bestimmt.
Für jedes δ(q,a,x) existiert kein entsprechendes δ(q,ε,x).
Die Bewegung von q_1 nach q_2 mit $\delta(q_1,\varepsilon,Z_0) = (q_2, Z_0)$ ist die einzige mögliche Bewegung bei einem leeren Speicher. Falls wir bei einem leeren Speicher noch Symbole der Eingabe lesen müssen, dann brechen wir einfach ab und das entsprechende Wort wird nicht akzeptiert.



<u>Gegenbeispiel:</u>

Die Sprache L_{wwr} kann nicht von einem deterministischen Kellerautomat akzeptiert werden => ist nichtdeterministisch. Daraus können wir schließen, dass die deterministischen kontextfreien Sprachen eine echte Untermenge der kontextfreien Sprachen sind.



3. Deterministische Kellerautomaten und reguläre Sprachen.

Deterministisch kontextfreie Sprachen umfassen eine Klasse von Sprachen, die zwischen den regulären Sprachen und den kontextfreien Sprachen liegt. Wir werden zuerst beweisen, dass die Sprachen von determinischen Kellerautomaten alle regulären Sprachen umfassen.

Wenn L eine reguläre Sprache ist, dann gibt es einen Kellerautomat P mit L = L(P).

<u>Beweis:</u>

Wir wissen, dass für jede reguläre Sprache kann man den entsprechenden endlichen Automat bauen, der diese Sprache erkennt. Also wir müssen zeigen, dass jeder endlichen Automat sich durch einen deterministischen Kellerautomat darstellen lässt.

Sei A = $(Q, \sum, \delta, q_0, F)$ ein deterministischen endlichen Automat. Wir konstruieren den entsprechenden Kellerautomat P folgender Weise:

$P = (Q, \sum, \Gamma, \delta, q_0, Z_0, F)$ mit:
$Q_p = Q_a$; $\sum_p = \sum_a$; $F_p = F_a$; $q_{0p} = q_{0a}$

Wir fügen Z_0 als Stackterminalsymbol, da ein Kellerautomat einen Stack besitzen muss. Daraus ist $\Gamma = \{Z_0\}$. Jetzt für alle q und p aus Q_a und a aus $\sum_a$ mit $\delta(q, a) = p$, definieren wir die entsprechende Übergangsfunktion in P: $\delta(q, a, Z_0) = (p, Z_0)$. Also wir führen den Zustandübergang, wobei wir den Stack nicht verändern.

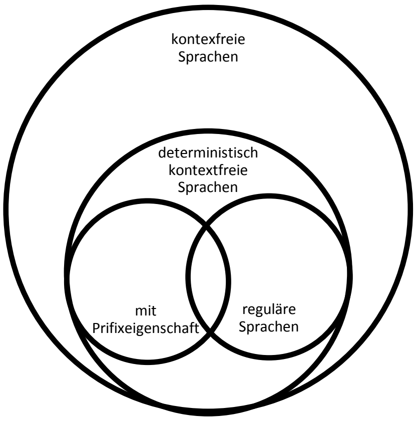
4. Deterministische Kellerautomaten mit Leerer-Keller-Akzeptanz

Wenn ein deterministischer Kellerautomat durch Leeren des Stacks akzeptieren soll, dann werden wir erkennen, dass wir über eine sehr beschränkte Fähigkeit zur Sprachenerkennung verfügen. Mit einem solchen Automat können wir nur kontextfreie Sprachen erkennen, die die Präfix-Eigenschaft haben. Das heißt, dass

wenn L eine Sprache, die von einem solchen Automat akzeptiert wird, für alle w aus L mit w=xy ist dann x kein Wort von L.

<u>Behauptung:</u>

Es gibt reguläre Sprachen, die von einem DKA mit Leerer-Keller-Akzeptanz nicht akzeptiert werden können (Beispiel: {0}*) und es gibt Sprachen die von einem DKA mit Leerer-Keller-Akzeptanz akzeptiert werden, aber nicht regulär sind (Beispiel – L_{wcwr}).



5. Deterministische Kellerautomaten und mehrdeutige Grammatiken

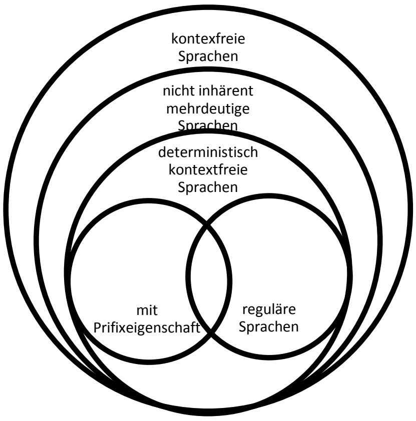
Sei L eine Sprache die eine eindeutige Grammatik (N,T,P,S) hat. Dann für jedes w aus L existiert nur ein Links- oder Rechtsableitungsbaum. Das heißt, dass für jedes w aus L genau eine Produktionsfolge $p_1,...,p_n$ mit p_i aus P existiert, die w generiert.

Wenn L eine Sprache ist, die von einem DKA akzeptiert wird, dann hat L eine eindeutige kontextfreie Grammatik.

Beweis:

Sei w ein beliebiges Wort aus L und K ein DKA, der L akzeptiert. Wenn K mit Eingabe w läuft, dann führt er eine deterministische Reihenfolge von Bewegungen um w zu akzeptieren. Wenn wir diese Folge kennen, dann können wir den Ableitungsbaum von w bestimmen und daraus die Produktionsfolge finden, die verwendet war um w zu konstruieren.

Also jede deterministisch kontextfreie Sprache hat eine eindeutige Grammatik, aber nicht jede Sprache, die eine eindeutige Grammatik hat, ist determnistisch kontextfrei(Bespiel: L_{wwr}).



6.Eigenschaften der deterministisch kontextfreien Sprachen.

6.1. abgeschlossen unter Komplement:

Sei L eine deterministisch kontextfreie Sprache -> existiert ein DKA K, der L akzeptiert. Wenn wir in K die akzeptierende und nicht akzeptierende Zustände tauschen, dann haben wir ein DKA K', der L^c akzeptiert -> L^c ist deterministisch kontextfrei.

6.2 nicht abgeschlossen unter Durchschnitt:

$L_1 = \{a^i b^j c^k | i=j\}$ -- deterministisch kontextfrei

$L_2 = \{a^i b^j c^k | j=k\}$ -- deterministisch kontextfrei

$L_1 \cap L_2 = \{a^n b^n c^n\}$ -- nicht kontextfrei

6.3 nicht abgeschlossen unter Vereinigung

$L_1 = \{a^i b^j c^k | i,j,k \geq 0 \text{ und } i \neq j\}$ -- deterministisch kontextfrei

$L_2 = \{a^i b^j c^k | i,j,k \geq 0 \text{ und } j \neq k\}$ -- deterministisch kontextfrei

$L = L_1 \cup L_2 = \{a^i b^j c^k | i,j,k \geq 0 \text{ und } (i \neq j) \text{ oder } (j \neq k)\}$

$L^c = \{a^i b^j c^k | i=j=k\} \cup \{(a,b,c)^* \mid \text{die Buchstaben sind in falscher Reihenfolge}\}$

$L' = L^c \cap a^*b^*c^* = \{a^n b^n c^n | n \geq 0\}$ -- nicht kontextfrei

Deterministisch kontextfreie Sprachen haben für die Praxis sehr nützliche Eigenschaft, dass für sie LR-Parser existieren, mit welchen in linearer Zeit beim Lesen von links nach rechts entschieden werden kann, ob die Eingabe ein Wort der Sprache ist. Viele in der Praxis verwendete formale Sprachen, insbesondere Programmiersprachen, gehören zu dieser Sprachklasse. In Vergleich dazu der schnellste Algorithmus, der das selbe Problem für nichtdeterministische kontextfreie Sprachen löst, läuft in $O(n^{2.378})$.

7.Quellen:

- E. Hopcroft, R. Motwani, J.D. Ullman. Einführung in die Automatentheorie, Formale Sprachen und Komplexität. Pearson Studium, 3. Auflage, 2011, Kapitel 6.4

http://www.grundstudium.info/theorie/node1.php

https://de.wikipedia.org/wiki/Kontextfreie_Grammatik

https://de.wikipedia.org/wiki/Mehrdeutige_Grammatik

https://de.wikipedia.org/wiki/Kellerautomat

https://en.wikipedia.org/wiki/LR_parser

http://www.cs.rice.edu/~yz2/COMP481/Chapt13.pdf

https://de.wikipedia.org/wiki/Deterministisch_kontextfreie_Sprache

https://cs.wmich.edu/~elise/courses/cs6800/DCFL.pptx

https://en.wikipedia.org/wiki/Pumping_lemma_for_context-free_languages