Chomsky-Grammatiken

- Ursprünglich von Chomsky in den 1950er Jahren eingeführt zur Beschreibung natürlicher Sprachen.
- Enge Verwandschaft zu Automaten
- Grundlage wichtiger Softwarekomponenten
- Enthalten außer den rechtslinearen und den kontextfreien weitere Grammatiktypen

Chomsky-Grammatik (Typ 0)

$G = (N, T, P, S)$ mit

- N: endl. Menge nichtterminaler Zeichen,
- T: endl. Menge terminaler Zeichen mit $N \cap T = \emptyset$,
- $P \subseteq (N \cup T)^* N (N \cup T)^* \times (N \cup T)^*$: endliche Menge von Produktionen
- $S \in N$: Startsymbol

▶ Schreibweise für Produktionen $(u, v) \in P$: $u ::= v$

Direkte Ableitung

$$w = xuy \underset{p}{\longrightarrow} xvy = w'$$

mit $w, w', x, y, u, v \in (N \cup T)^*$, $p = (u ::= v)$.

1. Suche u als Teilwort eines Wortes
2. Ersetze u durch v

- Schreibweise: $w \underset{P}{\longrightarrow} w'$,
  falls P eine Menge von Produktionen ist mit $p \in P$.

Ableitung (Iteration direkter Ableitungen)

$$w_0 \xrightarrow[p_1]{} w_1 \xrightarrow[p_2]{} \cdots \xrightarrow[p_n]{} w_n$$

für $w_0, \ldots, w_n \in (N \cup T)^*$ und Produktionen $p_1, \ldots, p_n$ $(n \geq 1)$

Schreibweisen:

- $w_0 \xrightarrow[P]{} \cdots \xrightarrow[P]{} w_n$ oder $w_0 \xrightarrow[P]{n} w_n$ oder $w_0 \xrightarrow[P]{*} w_n$, falls $p_1, \ldots, p_n \in P$.
- $w \xrightarrow{*} w'$, falls P aus dem Kontext klar ist.

Nullableitung

$$w \xrightarrow[P]{0} w$$

für alle $w \in (N \cup T)^*$.

Erzeugte Sprache

Sei $G = (N, T, P, S)$ eine Chomsky-Grammatik.

Erzeugte Sprache

$$L(G) = \{w \in T^* \mid S \xrightarrow[P]{*} w\}$$

Grammatiktypen

Sei $G = (N, T, P, S)$ eine Chomsky-Grammatik.

G heißt

- monoton (Typ 1), falls $length(u) \leq length(v)$
- kontextfrei (Typ 2), falls $length(u) = 1$
- regulär,rechtslinear (Typ 3), falls $length(u) = 1$ und $v \in T^* \cup T^+N$

für alle Produktionen $u ::= v$ aus P.

Chomsky-Hierarchie

Typ	Bezeichnung	Automaten
0	allgemein	Turing-Maschinen
1	monoton, kontext-sensitiv	linear beschränkte Automaten
2	kontextfrei	Kellerautomaten
3	regulär, rechtslinear	endliche Automaten

Wortprobleme

- Wortproblem von $L \subseteq T^*$: für alle $w \in T^*$ als Eingaben ist zu entscheiden, ob $w \in L$.
- Wortproblem $WP(G)$ einer Grammatik G ist Wortproblem von $L(G)$.
- $WP(G) \in O(n^3)$ für kontextfreie Grammatiken nach Cocke-Kasami-Younger.
- $WP(G) \in O(n)$, falls ein deterministischer Kellerautomat K existiert mit $L(G) = L(K)$.
- Insbes. $WP(G) \in O(n)$ für rechtslineares G wegen endlicher Automaten.

Typ	Bezeich.	Automaten	Wortproblem
0	allgemein	Turing-Maschinen	- (Reduktion Halteproblem auf WP)
1	kontext-sensitiv, monoton	linear beschränkte Automaten	$PSPACE = NPSPACE$
2	kontext-frei	Kellerautomaten	$O(n^3)$ (CKY)
3	regulär, rechtslinear	endliche Automaten	$O(n)$ (Det. endliche Automaten)

Wortproblem für kontextfreie Grammatiken

Das Wortproblem für kontextfreie Grammatiken ist mit einem Zeitaufwand der Größenordnung $O(n^3)$ lösbar.

- Cocke-Kasami-Younger-Algorithmus zur Lösung des Wortproblems kontextfreier Sprachen.
- Voraussetzung: Erzeugende Grammatik ist in Chomsky-Normalform.

Chomsky-Normalform (CNF)

- Vereinfacht die Lösung des Wortproblems für kontextfreie Sprachen.
- Vereinfacht den Beweis des Pumping-Lemmas für kontextfreie Sprachen.

Chomsky-Normalform (CNF)

Eine kontextfreie Grammatik $G = (N, T, P, S)$ ist in Chomsky-Normalform, falls für jede Regel $(A ::= r) \in P$ gilt:

$$r \in N^2 \text{ oder } r \in T$$

Satz

Für jede kontextfreie Grammatik G gibt es eine kontextfreie Grammatik G_{CNF} in Chomsky-Normalform, so dass

$$L(G) \setminus \{\lambda\} = L(G_{CNF}).$$

1.Schritt: Eliminierung von λ-Produktionen

Eine Produktion der Form

$$A ::= \lambda$$

heißt λ-Produktion.

Satz

Für jede kontextfreie Grammatik G gibt es eine kontextfreie Grammatik G' ohne λ-Produktionen, so dass

$$L(G) = L(G') \setminus \{\lambda\}.$$

Eliminierung von λ-Produktionen

Sei $G = (N, T, P, S)$ eine kontextfreie Grammatik.

1. (Sammeln aller $A \in N$ mit $A \xrightarrow{*} \lambda$.)

$$M_0 = \{A \in N \mid (A ::= \lambda) \in P\}$$

$$M_{i+1} = M_i \cup \{A \in N \mid (A ::= w) \in P, w \in M_i^*\}$$

Beobachtung

(a) Es existiert ein k, so dass $M_k = M_{k+1}$ gilt.

(b) Sei $k \in \mathbb{N}$ die kleinste Zahl mit $M_k = M_{k+1}$. Dann gilt:

$$A \xrightarrow[P]{*} \lambda \iff A \in M_k.$$

Eliminierung von λ-Produktionen

Sei $G = (N, T, P, S)$ eine kontextfreie Grammatik.

1. (Sammeln aller $A \in N$ mit $A \xrightarrow{*} \lambda$.)

$M_0 = \{A \in N \mid (A ::= \lambda) \in P\}$

$M_{i+1} = M_i \cup \{A \in N \mid (A ::= w) \in P, w \in M_i^*\}$

Beobachtung

(a) Es existiert ein k, so dass $M_k = M_{k+1}$ gilt.

(b) Sei $k \in \mathbb{N}$ die kleinste Zahl mit $M_k = M_{k+1}$. Dann gilt:

$$A \xrightarrow[P]{*} \lambda \iff A \in M_k.$$

2. (Konstruktion der neuen Regelmenge P')
   $P_0 = P$
   $P_{i+1} = P_i \cup \{A ::= u_1 u_2 \mid (A ::= u_1 B u_2) \in P_i, B \in M_k\}$
   Sei $m \in \mathbb{N}$ die kleinste Zahl mit $P_m = P_{m+1}$. Dann:
   $P' = P_m \setminus \{A ::= \lambda\}$
3. (Konstruktion der Grammatik G')

$$G' = (N, T, P', S)$$

Beispiel

$$\boxed{S ::= AB, A ::= aAA|\lambda, B ::= bBB|\lambda}$$

$$M_0 = \{A, B\},\ M_1 = \{A, B, S\},\ M_1 = M_2$$

$$P_0 :\ S ::= AB, A ::= aAA|\lambda, B ::= bBB|\lambda$$
$$P_1 :\ S ::= AB|A|B, A ::= aAA|aA|\lambda, B ::= bBB|bB|\lambda$$
$$P_2 :\ S ::= AB|A|B|\lambda, A ::= aAA|aA|a|\lambda, B ::= bBB|bB|b|\lambda$$
$$P_2 = P_3$$

$$P' :\ S ::= AB|A|B, A ::= aAA|aA|a, B ::= bBB|bB|b$$

2.Schritt: Eliminierung von Kettenregeln

Eine Produktion der Form

$$A ::= B$$

mit $B \in N$ heißt Kettenregel.

Satz

Für jede kontextfreie Grammatik G gibt es eine kontextfreie Grammatik G' ohne Kettenregeln, so dass

$$L(G) = L(G').$$

Eliminierung von Kettenregeln

1. (Finden aller Paare $(A, B) \in N \times N$ mit $A \xrightarrow{*} B$)
   $M_0 = \{(A, A) \mid A \in N\}$
   $M_{i+1} = M_i \cup \{(A, C) \mid (A, B) \in M_i, (B ::= C) \in P\}$

Beobachtung

Sei $k \in \mathbb{N}$ die kleinste Zahl mit $M_k = M_{k+1}$. Dann gilt:

$$(A, B) \in M_k \Longleftrightarrow A \xrightarrow{*} B.$$

2. (Konstruktion von G')

$$G' = (N, T, P', S)$$

mit

$$P' = \{A ::= r \mid (A, B) \in M_k, (B ::= r) \in P, r \notin N\}.$$

Beispiel

$$\boxed{S ::= A, A ::= B, B ::= b}$$

$$M_0 = \{(S,S),(A,A),(B,B)\}$$
$$M_1 = \{(S,S),(A,A),(B,B),(S,A),(A,B)\}$$
$$M_2 = \{(S,S),(A,A),(B,B),(S,A),(A,B),(S,B)\}$$

$$P' : \;\; S ::= b, A ::= b, B ::= b$$

3. und 4. Schritt

- Eliminieren der terminalen Symbole aus allen rechten Regelseiten der Länge ≥ 2 (4. Übungblatt).
- Verkürzen der rechten Regelseiten, deren Länge größer als 2 ist (vgl. Übung).

Eliminierung nutzloser Symbole

Ein Symbol $A \in N$ heißt nützlich, falls

$$S \xrightarrow[P]{*} u_1 A u_2 \xrightarrow[P]{*} w$$

mit $w \in T^*$.

Satz

Für jede kontextfreie Grammatik G gibt es eine kontextfreie Grammatik G' ohne nutzlose Symbole, so dass

$$L(G) = L(G').$$