

Kanonické tvary bezkontextových gramatik

- redukované bezkontextové gramatiky
- gramatiky bez ε -pravidel
- gramatiky bez jednoduchých pravidel
- vlastní gramatiky
- Chomského normální forma
- gramatiky bez levé rekurze
- Greibachové normální forma

Rekurzivní neterminály a gramatiky

Definice 3.28. Neterminál A v CFG $\mathcal{G} = (N, \Sigma, P, S)$ se nazývá **levorekurzivní** jestliže v \mathcal{G} existuje derivace $A \Rightarrow^+ A\beta$.

CFG bez levorekurzivních neterminálů se nazývá **nelevorekurzivní**.

Je-li v CFG pravidlo tvaru $A \rightarrow A\alpha$, hovoříme o **přímé levé rekurzi** na A .

Praktický význam: některé nástroje pro automatickou tvorbu parserů k zadaným gramatikám vyžadují na vstupu nelevorekurzivní gramatiku (např. ANTLR).

Algoritmus odstranění přímé levé rekurze

Nechť CFG $\mathcal{G} = (N, \Sigma, P, S)$ je necyklická a bez ε -pravidel, v níž všechna A -pravidla (pravidla mající na levé straně A) jsou tvaru

$$A \rightarrow A\alpha_1 \mid \dots \mid A\alpha_m \mid \beta_1 \mid \dots \mid \beta_n,$$

kde každý řetěz β_i začíná symbolem různým od A .

Nechť $\mathcal{G}' = (N \cup \{A'\}, \Sigma, P', S)$, kde P' obdržíme z P tak, že všechna výše uvedená pravidla nahradíme pravidly:

$$A \rightarrow \beta_1 \mid \dots \mid \beta_n \mid \beta_1 A' \mid \dots \mid \beta_n A'$$

$$A' \rightarrow \alpha_1 \mid \dots \mid \alpha_m \mid \alpha_1 A' \mid \dots \mid \alpha_m A'$$

Pak $L(\mathcal{G}) = L(\mathcal{G}')$ a \mathcal{G}' je necyklická a bez ε -pravidel.

Lemma o substituci

Lemma 3.20. (o substituci)

Nechť $\mathcal{G} = (N, \Sigma, P, S)$ je CFG. Nechť $A \rightarrow \alpha_1 B \alpha_2 \in P$.

Nechť $B \rightarrow \beta_1 \mid \dots \mid \beta_r$ jsou všechna pravidla v P tvaru $B \rightarrow \alpha$.

Definujme $\mathcal{G}' = (N, \Sigma, P', S)$, kde

$$P' = (P \setminus \{A \rightarrow \alpha_1 B \alpha_2\}) \cup \{A \rightarrow \alpha_1 \beta_1 \alpha_2 \mid \dots \mid \alpha_1 \beta_r \alpha_2\}.$$

Pak $L(\mathcal{G}) = L(\mathcal{G}')$.

Příklad

$A \rightarrow Bd \mid c$

$B \rightarrow Bdd \mid Ccc \mid aAd$

$C \rightarrow Aa$

Algoritmus odstranění levé rekurze

Vstup: Vlastní CFG $\mathcal{G} = (N, \Sigma, P, S)$

Výstup: Ekvivalentní nelevorekurzivní gramatika bez ε -pravidel

- 1: Uspořádej libovolně N , $N = \{A_1, \dots, A_n\}$
- 2: **for** $i \leftarrow 1$ **to** n **do**
- 3: **for** $j \leftarrow 1$ **to** $i - 1$ **do**
- 4: **for all** pravidlo tvaru $A_i \rightarrow A_j\alpha$ **do**
- 5: přidej pravidla $A_i \rightarrow \beta_1\alpha \mid \dots \mid \beta_k\alpha$
- 6: (kde $A_j \rightarrow \beta_1 \mid \dots \mid \beta_k$ jsou všechna pravidla pro A_j)
- 7: vypuť pravidlo $A_i \rightarrow A_j\alpha$
- 8: **end for**
- 9: **end for**
- 10: odstraň případnou přímou levou rekurzi na A_i
- 11: **end for**

Korektnost algoritmu

Konečnost.

Ekvivalence gramatik: Všechny úpravy jsou dle Lemmatu o substituci nebo odstraňují přímou levou rekurzi.

Výsledná gramatika je nelevorekurzivní:

- 1 po i -té iteraci vnějšího cyklu začíná každé A_i -pravidlo buď terminálem nebo neterminálem A_k , kde $k > i$.
- 2 po j -té iteraci vnitřního cyklu začíná každé A_i -pravidlo buď terminálem nebo neterminálem A_k , kde $k > j$.

Výsledná gramatika je bez ε -pravidel.

Greibachové normální forma

Definice 3.33. Bezkontextová gramatika $\mathcal{G} = (N, \Sigma, P, S)$ je v **Greibachové normální formě** (GNF), právě když

- \mathcal{G} je bez ε -pravidel a
- každé pravidlo z P je tvaru $A \rightarrow a\alpha$, kde $a \in \Sigma$ a $\alpha \in N^*$ (s případnou výjimkou pravidla $S \rightarrow \varepsilon$).

Věta 3.34. Každý bezkontextový jazyk lze generovat bezkontextovou gramatikou v Greibachové normální formě.

Důkaz.

$L = \emptyset$: zřejmé ($S \rightarrow aS$)

$L \neq \emptyset$: 1. z vlastní gramatiky eliminujeme levou rekurzi
2. pak převedeme do GNF



Příklad

$A \rightarrow Ba \mid Db \mid c$

$B \rightarrow CC$

$C \rightarrow aE$

$D \rightarrow CDa \mid Eb$

$E \rightarrow bb$

Algoritmus transformace do GNF

Vstup: Nelevorekurzivní CFG $\mathcal{G} = (N, \Sigma, P, S)$ bez ε -pravidel

Výstup: Ekvivalentní gramatika v GNF

- 1: Najdi lineární uspořádání \prec splňující $(A \rightarrow B\alpha) \in P \implies A \prec B$
- 2: Označme $N = \{A_1, \dots, A_n \mid A_{i-1} \prec A_i, 1 < i \leq n\}$
- 3: **for** $i \leftarrow n - 1$ **downto** 1 **do**
- 4: **for all** pravidlo tvaru $A_i \rightarrow A_j\alpha$, kde $j > i$ **do**
- 5: přidej pravidlo $A_i \rightarrow \beta_1\alpha \mid \dots \mid \beta_k\alpha$
- 6: (kde $A_j \rightarrow \beta_1 \mid \dots \mid \beta_k$ jsou všechna A_j -pravidla)
- 7: vypuť pravidlo $A_i \rightarrow A_j\alpha$
- 8: **end for**
- 9: **end for**
- 10: Nahrad' potřebné terminály novými neterminály
- 11: a přidej příslušná pravidla

Korektnost algoritmu

Konečnost.

Ekvivalence gramatik.

Výsledná gramatika je v GNF.

Zásobníkové automaty

Definice zásobníkového automatu

Definice 3.36. Nedeterministický zásobníkový automat (PushDown Automaton, PDA) je sedmice $\mathcal{M} = (Q, \Sigma, \Gamma, \delta, q_0, Z_0, F)$, kde

- Q je konečná množina, jejíž prvky nazýváme **stavy**,
- Σ je konečná množina, tzv. **vstupní abeceda**,
- Γ je konečná množina, tzv. **zásobníková abeceda**,
- $\delta : Q \times (\Sigma \cup \{\varepsilon\}) \times \Gamma \rightarrow \mathcal{P}_{Fin}(Q \times \Gamma^*)$, tzv. (parciální) **přechodová funkce**¹,
- $q_0 \in Q$ je **počáteční stav**,
- $Z_0 \in \Gamma$ je **počáteční symbol v zásobníku**,
- $F \subseteq Q$ je množina **koncových stavů**.

¹Zápis $\mathcal{P}_{Fin}(Q \times \Gamma^*)$ značí množinu všech **konečných** podmnožin množiny $Q \times \Gamma^*$.

Výpočet zásobníkového automatu

Definice 3.37. Necht' $\mathcal{M} = (Q, \Sigma, \Gamma, \delta, q_0, Z_0, F)$ je PDA.

Konfigurací nazveme libovolný prvek $(p, w, \alpha) \in Q \times \Sigma^* \times \Gamma^*$.

Na množině všech konfigurací automatu \mathcal{M} definujeme binární relaci

krok výpočtu $\vdash_{\mathcal{M}}$ takto:

$$(p, aw, Z\alpha) \vdash_{\mathcal{M}} (q, w, \gamma\alpha) \stackrel{\text{def}}{\iff} \exists (q, \gamma) \in \delta(p, a, Z) \text{ pro } a \in \Sigma \cup \{\varepsilon\}$$

Reflexivní a tranzitivní uzávěr relace $\vdash_{\mathcal{M}}$ značíme $\vdash_{\mathcal{M}}^*$.

Je-li \mathcal{M} zřejmý z kontextu, píšeme pouze \vdash resp. \vdash^* .

Akceptující výpočet zásobníkového automatu

Definice 3.37. (pokračování)

Jazyk akceptovaný PDA \mathcal{M} koncovým stavem definujeme jako

$$L(\mathcal{M}) = \{w \in \Sigma^* \mid (q_0, w, Z_0) \xrightarrow{*} (q_f, \varepsilon, \alpha), \text{ kde } q_f \in F, \alpha \in \Gamma^*\}$$

a jazyk akceptovaný PDA \mathcal{M} **prázdným zásobníkem** definujeme jako

$$L_e(\mathcal{M}) = \{w \in \Sigma^* \mid (q_0, w, Z_0) \xrightarrow{*} (q, \varepsilon, \varepsilon), \text{ kde } q \in Q\}.$$