

Kontekstno-neodvisne gramatike za kodiranje in stiskanje podatkov

Janez Podlogar

Univerza v Ljubljani, Fakulteta za matematiko in fiziko

16. 9. 2024

- Kodiranje je spreminjanje zapisa sporočila.
- Stiskanje podatkov je zapis sporočilo v zgoščeni obliki.
- Gramatike je nabor pravil za tvorjenje nizov.

Kontekstno-neodvisna gramatika

Definicija

Kontekstno-neodvisna gramatika je četverica $G = (V, \Sigma, P, S)$, kjer je

- V abeceda *nekončnih simbolov*;
- Σ abeceda *končnih simbolov* taka, da $\Sigma \cap V = \emptyset$;
- $P \subseteq V \times (V \cup \Sigma)^*$ celovita relacija;
- $S \in V$ *začetni simbol*.

Kontekstno-neodvisna gramatika

Definicija

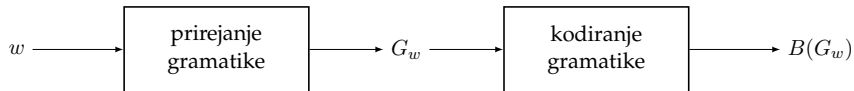
Kontekstno-neodvisna gramatika je četverica $G = (V, \Sigma, P, S)$, kjer je

- V abeceda *nekončnih simbolov*;
- Σ abeceda *končnih simbolov* taka, da $\Sigma \cap V = \emptyset$;
- $P \subseteq V \times (V \cup \Sigma)^*$ celovita relacija;
- $S \in V$ *začetni simbol*.

Definicija

Jezik KNG je množica nizov, ki jih lahko izpeljemo iz začetnega simbola in ne vsebujejo nekončnih simbolov.

Stiskanje niza



Dopustna gramatika

Definicija

KNG je *deterministična*, če vsak nekončen simbol $A \in V$ nastopa natanko enkrat kot leva stran nekega prepisovalnega pravila.

Dopustna gramatika

Definicija

KNG je *deterministična*, če vsak nekončen simbol $A \in V$ nastopa natanko enkrat kot leva stran nekega prepisovalnega pravila.

Trditev

Jezik deterministične KNG je enojec ali prazna množica.

Definicija

KNG *ne vsebuje neuporabnih simbolov*, če se vsak simbol $y \in V \cup \Sigma$, $y \neq S$ pojavi vsaj enkrat v izpeljavi niza, ki je v jeziku KNG.

Definicija

KNG je *dopustna gramatika*, če je:

- deterministična,
- ne vsebuje neuporabnih simbolov,
- ima neprazen jezik,
- prazen niz ne nastopa kot desna stran kateregakoli prepisovalnega pravila.

Definicija

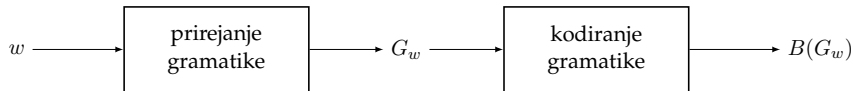
KNG je *dopustna gramatika*, če je:

- deterministična,
- ne vsebuje neuporabnih simbolov,
- ima neprazen jezik,
- prazen niz ne nastopa kot desna stran kateregakoli prepisovalnega pravila.

Posledica

Jezik dopustne gramatike je enojec.

Prirejanje gramatike



Asimptotsko kompaktno prirejanje gramatike

Definicija

Prirejanje gramatike nizu abecede \mathcal{A} je *asimptotsko kompaktno*, če za vsak niz $w \in \mathcal{A}^+$ velja

$$\lim_{n \rightarrow \infty} \max_{w \in \mathcal{A}^n} \frac{|G_w|}{|w|} = 0.$$

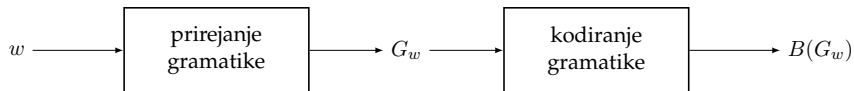
Neskrčljivo prirejanje gramatike

Definicija

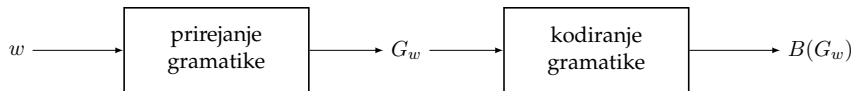
Pravimo, da je KNG G *neskrčljiva gramatika*, če:

- 1 Vsak $A \in V$, $A \neq S$ nastopa vsaj dvakrat v desni strani prepisovalnih pravil;
- 2 Ne obstajata $y_1, y_2 \in V \cup \Sigma$, da niz y_1y_2 nastopa kot podniz desne strani kateregali prepisovalnega pravila več kot enkrat na neprekrivajočih se mestih.

Binarno kodiranje gramatike



Binarno kodiranje gramatike



Izrek

Obstaja bijektivno brezpredponsko binarno kodiranje gramatike, da

$$\forall G \in \mathcal{G}(\mathcal{A}): |B(G)| \leq |\mathcal{A}| + 4|G| + \lceil H(G) \rceil.$$

Definicija

Stiskanje niza abecede \mathcal{A} z gramatikami je par preslikav kodne in dekodne preslikave $\Phi = (\kappa, \delta)$. Kodna preslikava je

$$\begin{aligned}\kappa: \mathcal{A}^+ &\rightarrow \{0, 1\}^+, \\ w &\mapsto B(\pi(w)),\end{aligned}$$

kjer je π prirejanje gramatike nizu abecede \mathcal{A} in B binarno kodiranje dopustne gramatike.