Kontekstno-neodvisne gramatike za kodiranje in stiskanje podatkov

Janez Podlogar

Univerza v Ljubljani, Fakulteta za matematiko in fiziko

5. 5. 2023

Spomnimo se od prejšnjič

S kontekstno-neodivsnih gramatik želimo stisniti poljuben niz.

Spomnimo se od prejšnjič

S kontekstno-neodivsnih gramatik želimo stisniti poljuben niz.

Definicija

- *Abeceda* je končna neprazna množica Σ
- *Množica vseh končnih nizov abecede* Σ označimo s Σ^*
- *Dolžina niza* w je število znakov abecede v w, označimo z $|w|$

Spomnimo se od prejšnjič

S kontekstno-neodivsnih gramatik želimo stisniti poljuben niz.

Definicija

- *Abeceda* je končna neprazna množica Σ
- *Množica vseh končnih nizov abecede* Σ označimo s Σ^*
- *Dolžina niza* w je število znakov abecede v w, označimo z $|w|$

Primer nizov abecede

Naj bo $\Sigma = \{a, b, c\}$ abeceda, potem sta niza

$$ab \in \Sigma^*, \quad cababcccababcccab \in \Sigma^*$$

Kontekstno-neodvisne gramatike

Definicija

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

Kontekstno-neodvisne gramatike

Definicija

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

- V končna množica *nekončnih simbolov*

Kontekstno-neodvisne gramatike

Definicija

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

- V končna množica *nekončnih simbolov*
- abeceda Σ množica *končnih simbolov*

Kontekstno-neodvisne gramatike

Definicija

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

- V končna množica *nekončnih simbolov*
- abeceda Σ množica *končnih simbolov*
- $P \subseteq V \times (V \cup \Sigma)^*$ celovita relacija

Kontekstno-neodvisne gramatike

Definicija

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

- V končna množica *nekončnih simbolov*
- abeceda Σ množica *končnih simbolov*
- $P \subseteq V \times (V \cup \Sigma)^*$ celovita relacija
- $S \in V$ *začetni simbol*

Kontekstno-neodvisne gramatike

Definicija

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

- V končna množica *nekončnih simbolov*
- abeceda Σ množica *končnih simbolov*
- $P \subseteq V \times (V \cup \Sigma)^*$ celovita relacija
- $S \in V$ *začetni simbol*

Definicija

Jezik kontekstno-neodvisne gramatike G je množica vseh nizov, ki jih lahko izpeljemo z gramatiko G, označimo ga z $L(G)$.

Stiskanja niza s kontekstno-neodvisno gramatiko

Stiskanje niza $w = cababcccababcccab$

Naj bo $G_w = (V, \Sigma, P, S)$, kjer je

Stiskanja niza s kontekstno-neodvisno gramatiko

Stiskanje niza $w = cababcccababcccab$

Naj bo $G_w = (V, \Sigma, P, S)$, kjer je

- $V = \{S, A, B, C\}$

Stiskanja niza s kontekstno-neodvisno gramatiko

Stiskanje niza $w = cababcccababcccab$

Naj bo $G_w = (V, \Sigma, P, S)$, kjer je

- $V = \{S, A, B, C\}$
- $\Sigma = \{a, b, c\}$

Stiskanja niza s kontekstno-neodvisno gramatiko

Stiskanje niza $w = cababcccababcccab$

Naj bo $G_w = (V, \Sigma, P, S)$, kjer je

- $V = \{S, A, B, C\}$
- $\Sigma = \{a, b, c\}$
- $P = \{S \to cCCA, A \to ab, B \to ccc, C \to AAB\}$

Stiskanja niza s kontekstno-neodvisno gramatiko

Stiskanje niza $w = cababcccababcccab$

Naj bo $G_w = (V, \Sigma, P, S)$, kjer je

- $V = \{S, A, B, C\}$
- $\Sigma = \{a, b, c\}$
- $P = \{S \to cCCA, A \to ab, B \to ccc, C \to AAB\}$
- $S = S$

Stiskanja niza s kontekstno-neodvisno gramatiko

Stiskanje niza $w = cababcccababcccab$

Naj bo $G_w = (V, \Sigma, P, S)$, kjer je

- $V = \{S, A, B, C\}$
- $\Sigma = \{a, b, c\}$
- $P = \{S \to cCCA, A \to ab, B \to ccc, C \to AAB\}$
- $S = S$

$S = cCCA$

Stiskanja niza s kontekstno-neodvisno gramatiko

Stiskanje niza $w = cababcccababcccab$

Naj bo $G_w = (V, \Sigma, P, S)$, kjer je

- $V = \{S, A, B, C\}$
- $\Sigma = \{a, b, c\}$
- $P = \{S \to cCCA, A \to ab, B \to ccc, C \to AAB\}$
- $S = S$

$S = cCCA \overset{C}{\Rightarrow} cAABAAB$

Stiskanja niza s kontekstno-neodvisno gramatiko

Stiskanje niza $w = cababcccababcccab$

Naj bo $G_w = (V, \Sigma, P, S)$, kjer je

- $V = \{S, A, B, C\}$
- $\Sigma = \{a, b, c\}$
- $P = \{S \to cCCA, A \to ab, B \to ccc, C \to AAB\}$
- $S = S$

$S = cCCA \overset{C}{\Rightarrow} cAABAAB \overset{B}{\Rightarrow} cAAcccAAccc$

Stiskanja niza s kontekstno-neodvisno gramatiko

Stiskanje niza $w = cababcccababcccab$

Naj bo $G_w = (V, \Sigma, P, S)$, kjer je

- $V = \{S, A, B, C\}$
- $\Sigma = \{a, b, c\}$
- $P = \{S \to cCCA, A \to ab, B \to ccc, C \to AAB\}$
- $S = S$

$S = cCCA \overset{C}{\Rightarrow} cAABAAB \overset{B}{\Rightarrow} cAAcccAAccc \overset{A}{\Rightarrow} cababcccababccc$

Stiskanja niza s kontekstno-neodvisno gramatiko

Stiskanje niza $w = cababcccababcccab$

Naj bo $G_w = (V, \Sigma, P, S)$, kjer je

- $V = \{S, A, B, C\}$
- $\Sigma = \{a, b, c\}$
- $P = \{S \to cCCA, A \to ab, B \to ccc, C \to AAB\}$
- $S = S$

$$S = cCCA \overset{C}{\Longrightarrow} cAABAAB \overset{B}{\Longrightarrow} cAAcccAAccc \overset{A}{\Longrightarrow} cababcccababccc = w$$

Stiskanja niza s kontekstno-neodvisno gramatiko

Stiskanje niza $w = cababcccababcccab$

Naj bo $G_w = (V, \Sigma, P, S)$, kjer je

- $V = \{S, A, B, C\}$
- $\Sigma = \{a, b, c\}$
- $P = \{S \to cCCA, A \to ab, B \to ccc, C \to AAB\}$
- $S = S$

$S = cCCA \overset{C}{\Rightarrow} cAABAAB \overset{B}{\Rightarrow} cAAcccAAccc \overset{A}{\Rightarrow} cababcccababccc = w$

Velja, da je

$$L(G_w) = \{w\}.$$

Deterministična kontekstno-neodvisna gramatika

Definicija

Kontekstno-neodvisna gramatika G je *deterministična*, če vsak nekončen simbol $A \in V$, nastopa natanko enkrat kot levi član nekega produkcijskega pravila. Kontekstno-neodvisna gramatika, ki ni deterministična, je *nedeterministična*.

Deterministična kontekstno-neodvisna gramatika

Definicija

Kontekstno-neodvisna gramatika G je *deterministična*, če vsak nekončen simbol $A \in V$, nastopa natanko enkrat kot levi člen nekega produkcijskega pravila. Kontekstno-neodvisna gramatika, ki ni deterministična, je *nedeterministična*.

Trditev

Naj bo G deterministična kontekstno-neodvisna gramatika. Potem je jezik gramatike G enojec ali pa prazna množica.

Deterministična kontekstno-neodvisna gramatika

Deterministična kontekstno-neodvisna gramatika z praznim jeziokm

Naj bo $G = (V, \Sigma, P, S)$, kjer je

Deterministična kontekstno-neodvisna gramatika

Deterministična kontekstno-neodvisna gramatika z praznim jeziokm

Naj bo $G = (V, \Sigma, P, S)$, kjer je

- $V = \{S\}$

Deterministična kontekstno-neodvisna gramatika

Deterministična kontekstno-neodvisna gramatika z praznim jeziokm

Naj bo $G = (V, \Sigma, P, S)$, kjer je

- $V = \{S\}$
- $\Sigma = \{a\}$

Deterministična kontekstno-neodvisna gramatika

Deterministična kontekstno-neodvisna gramatika z praznim jeziokm

Naj bo $G = (V, \Sigma, P, S)$, kjer je

- $V = \{S\}$
- $\Sigma = \{a\}$
- $P = \{S \to S\}$

Deterministična kontekstno-neodvisna gramatika

Deterministična kontekstno-neodvisna gramatika z praznim jeziokm

Naj bo $G = (V, \Sigma, P, S)$, kjer je

- $V = \{S\}$
- $\Sigma = \{a\}$
- $P = \{S \to S\}$
- $S = S$

Deterministična kontekstno-neodvisna gramatika

Deterministična kontekstno-neodvisna gramatika z praznim jeziokm

Naj bo $G = (V, \Sigma, P, S)$, kjer je

- $V = \{S\}$
- $\Sigma = \{a\}$
- $P = \{S \to S\}$
- $S = S$

Gramatika je očitno deterministična in velja $L(G) = \emptyset$.

Deterministična kontekstno-neodvisna gramatika

Deterministična kontekstno-neodvisna gramatika z odvečnimi simboli

Naj bo $G = (V, \Sigma, P, S)$, kjer je

- $V = \{S, A\}$
- $\Sigma = \{a\}$
- $P = \{S \to a, A \to a\}$
- $S = S$

Deterministična kontekstno-neodvisna gramatika

Deterministična kontekstno-neodvisna gramatika z odvečnimi simboli

Naj bo $G = (V, \Sigma, P, S)$, kjer je

- $V = \{S, A\}$
- $\Sigma = \{a\}$
- $P = \{S \to a, A \to a\}$
- $S = S$

Gramatika je očitno deterministična in velja $L(G) = a$.

Neuporabni simboli in prazen jezik

Definicija

Pravimo, da kontekstno-neodvisna gramatika G *ne vsebuje neuporabnih simbolov*, ko za vsak simbol $y \in V \cup \Sigma,\ y \neq S$, obstaja končno mnogo nizov $\alpha_1, \alpha_2, \ldots, \alpha_n \in (V \cup \Sigma)^*$ tako, da je y vsebovan vsaj v enem izmed nizov in velja

$$S = \alpha_1 \Rightarrow \alpha_2 \Rightarrow \cdots \alpha_n \in L(G).$$

Dopustne gramatike

Definicija

Kontekstno-neodvisna gramatika G je *dopustna*, če je deterministična, ne vsebuje neuporabnih simbolov, $L(G) \neq \emptyset$ in prazen niz ne nastopa kot desni član kateregakoli produkcijska pravila v P.

Dopustne gramatike

Definicija

Kontekstno-neodvisna gramatika G je *dopustna*, če je deterministična, ne vsebuje neuporabnih simbolov, $L(G) \neq \emptyset$ in prazen niz ne nastopa kot desni član kateregakoli produkcijska pravila v P.

Posledica

Jezik dopustne kontekstno-neodvisne gramatike je enojec.

Za dopustno gramatiko je dovolj podati produkcijska pravila

Podana imamo produckcijska pravila

- $P = \{A_0 \rightarrow aA_1A_2A_3, A_1 \rightarrow ab, A_2 \rightarrow A_1b, A_3 \rightarrow A_2b\}$

Za dopustno gramatiko je dovolj podati produkcijska pravila

Podana imamo produckcijska pravila

- $P = \{A_0 \to aA_1A_2A_3, A_1 \to ab, A_2 \to A_1b, A_3 \to A_2b\}$

Nekončni simboli gramatike so

$$V = \{A_0, A_1, A_2, A_3\}.$$

Za dopustno gramatiko je dovolj podati produkcijska pravila

Podana imamo produckcijska pravila

- $P = \{A_0 \to aA_1A_2A_3, A_1 \to ab, A_2 \to A_1b, A_3 \to A_2b\}$

Nekončni simboli gramatike so

$$V = \{A_0, A_1, A_2, A_3\}.$$

Končni simboli gramatike so

$$\Sigma = \{a, b\}.$$

Za dopustno gramatiko je dovolj podati produkcijska pravila

Podana imamo produckcijska pravila

- $P = \{A_0 \to aA_1A_2A_3, A_1 \to ab, A_2 \to A_1b, A_3 \to A_2b\}$

Nekončni simboli gramatike so

$$V = \{A_0, A_1, A_2, A_3\}.$$

Končni simboli gramatike so

$$\Sigma = \{a, b\}.$$

Začetni simbol je

$$S = A_0.$$

Endomorfiem

Definicija

Naj bo Σ abeceda. *Endomorfizem množice* Σ^* je preslikava $f\colon \Sigma^* \to \Sigma^*$ tako, da je

$$f(\varepsilon) = \varepsilon \text{ in } \forall w, u \in \Sigma^* : f(wu) = f(w)f(u).$$

Endomorfiem

Definicija

Naj bo Σ abeceda. *Endomorfizem množice* Σ^* je preslikava $f\colon \Sigma^* \to \Sigma^*$ tako, da je

$$f(\varepsilon) = \varepsilon \text{ in } \forall w, u \in \Sigma^* : f(wu) = f(w)f(u).$$

Za endomorfizem f množice Σ^* induktivno definiramo

$$\begin{aligned} f^0(w) &= w, \\ f^1(w) &= f(w), \\ f^k(w) &= f(f^{k-1}(w)), \end{aligned}$$

kjer je $w \in \Sigma^*$ in $k \geq 2$ celo število.

D0L sistem

Definicija

D0L sistem je trojica $D = (\Sigma, f, w)$, kjer je

D0L sistem

Definicija

D0L sistem je trojica $D = (\Sigma, f, w)$, kjer je

- Σ abeceda

D0L sistem

Definicija

D0L sistem je trojica $D = (\Sigma, f, w)$, kjer je

- Σ abeceda
- f endomorfizem množice Σ^*

D0L sistem

Definicija

D0L sistem je trojica $D = (\Sigma, f, w)$, kjer je

- Σ abeceda
- f endomorfizem množice Σ^*
- $w \in \Sigma^*$ *aksiom*

D0L sistem

Definicija

D0L sistem je trojica $D = (\Sigma, f, w)$, kjer je

- Σ abeceda
- f endomorfizem množice Σ^*
- $w \in \Sigma^*$ *aksiom*

Sistem generira zaporedje nizov $\{f^k(w) \mid k = 0, 1, 2 \ldots\}$, ki ima *fiksno točko* w^*, če velja

$$w^* \in \{f^k(w) \mid k = 0, 1, 2 \ldots\},$$
$$f(w^*) = w^*.$$

D0L sistem prirejen gramatiki

Definicija

Naj bo G deterministična kontekstno-neodvisna gramatika v kateri prazen niz ne nastopa kot desni član kateregakoli produkcijska pravila. Na $(V \cup \Sigma)^*$ definiramo endomorfizem f_G tako, da

$$\forall a \in \Sigma \colon f_G(a) = a;$$

$$\text{če je } A \to \alpha \text{ produkcijsko pravilo, potem je } f_G(A) = \alpha.$$

D0L sistem prirejen gramatiki

Definicija

Naj bo G deterministična kontekstno-neodvisna gramatika v kateri prazen niz ne nastopa kot desni član kateregakoli produkcijska pravila. Na $(V \cup \Sigma)^*$ definiramo endomorfizem f_G tako, da

$$\forall a \in \Sigma \colon f_G(a) = a;$$

$$\text{če je } A \to \alpha \text{ produkcijsko pravilo, potem je } f_G(A) = \alpha.$$

D0L sistem $(V \cup \Sigma,\ f_G,\ S)$ označimo z $D(G)$ in ga imenujemo *D0L sistem prirejen gramatiki G*.

Karakterizacija dopustnih gramatik preko D0L sistema

Izrek

Naj bo G dopustna kontekstno-neodvisna gramatika. Potem jezik gramatike G ustreza fiksni točki D0L sistema prirejenega gramatiki G.

Karakterizacija dopustnih gramatik preko D0L sistema

Izrek

Naj bo G dopustna kontekstno-neodvisna gramatika. Potem jezik gramatike G ustreza fiksni točki D0L sistema prirejenega gramatiki G.

Izrek

Naj bo G kontekstno-neodvisna gramatika v kateri prazen niz ne nastopa kot desni član kateregakoli produkcijska pravila. Gramatika G je dopustna natanko takrat, ko je $f_G^{|V|}(S) \in \Sigma^+$ in vsak simbol iz $V \cup \Sigma$ nastopa v vsaj enem izmed nizov $f_G^i(S)$, kjer je $i = 0, 1, \ldots, |V|$.