Kontekstno-neodvisne gramatike za kodiranje in stiskanje podatkov

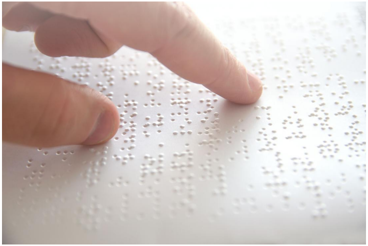
Janez Podlogar

Univerza v Ljubljani, Fakulteta za matematiko in fiziko

21. 11. 2022

Kodiranje in kod

Slika: Telegrafska tipka in zvočnik.



Slika: Braillova pisava.

Kodiranje in kod

- Spreminjanje zapisa sporočila imenujemo *kodiranje*.

Kodiranje in kod

- Spreminjanje zapisa sporočila imenujemo *kodiranje*.
- Sistem pravil po katerem se kodiranje opravi imenujemo *kod*.

Morsejeva abeceda

Morsejeva abeceda je kodiranje črk, števil in ločil s pomočjo zaporedja signalov:

Morsejeva abeceda

Morsejeva abeceda je kodiranje črk, števil in ločil s pomočjo zaporedja signalov:

- Dolžina kratkega signala je ena enota.

Morsejeva abeceda

Morsejeva abeceda je kodiranje črk, števil in ločil s pomočjo zaporedja signalov:

- Dolžina kratkega signala je ena enota.
- Dolgi signal je trikrat daljši od kratkega signala.

Morsejeva abeceda

Morsejeva abeceda je kodiranje črk, števil in ločil s pomočjo zaporedja signalov:

- Dolžina kratkega signala je ena enota.
- Dolgi signal je trikrat daljši od kratkega signala.
- Razmik med signali znotraj črke je tišina dolžine kratkega signala.

Morsejeva abeceda

Morsejeva abeceda je kodiranje črk, števil in ločil s pomočjo zaporedja signalov:

- Dolžina kratkega signala je ena enota.
- Dolgi signal je trikrat daljši od kratkega signala.
- Razmik med signali znotraj črke je tišina dolžine kratkega signala.
- Razmik med črkami je tišina dolga tri kratke signale oz. en dolgi signal.

Morsejeva abeceda

Morsejeva abeceda je kodiranje črk, števil in ločil s pomočjo zaporedja signalov:

- Dolžina kratkega signala je ena enota.
- Dolgi signal je trikrat daljši od kratkega signala.
- Razmik med signali znotraj črke je tišina dolžine kratkega signala.
- Razmik med črkami je tišina dolga tri kratke signale oz. en dolgi signal.
- Presledek med besedami je tišina dolga sedmih kratkih signalov.

Abeceda in nizi na abecedi

Definicija

- *Abeceda* je končna neprazna množica Σ.
- *Množica vseh končnih nizov abecede* Σ označimo z Σ^*.

Abeceda in nizi na abecedi

Definicija

- *Abeceda* je končna neprazna množica Σ.
- *Množica vseh končnih nizov abecede* Σ označimo z Σ^*.

Primer nizov abecede

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

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

Kodiranje in dekodiranje

Definicija

- *Kodiranje nizov abecede* Σ je injektivna funkcija $\kappa\colon \Sigma^* \to \Sigma_c^*$.

Kodiranje in dekodiranje

Definicija

- *Kodiranje nizov abecede* Σ je injektivna funkcija $\kappa\colon \Sigma^* \to \Sigma_c^*$.
- *Dokodiranje kodiranja* κ je funkcija $\delta\colon C \subseteq \Sigma_c^* \to \Sigma^*$, da velja

$$\forall w \in \Sigma^*\colon \delta(\kappa(w)) = w.$$

Morsejeva abeceda

Morsov kod

Morsejeva abeceda

Morsov kod

- $\Sigma = \{A, B, \ldots, Z\} \cup \{0, 1, \ldots, 9\} \cup \{\textvisiblespace\}$

Morsejeva abeceda

Morsov kod

- $\Sigma = \{\mathrm{A}, \mathrm{B}, \ldots, \mathrm{Z}\} \cup \{0, 1, \ldots, 9\} \cup \{\textvisiblespace\}$
- $\Sigma_c = \{\cdot, -, \square\}$

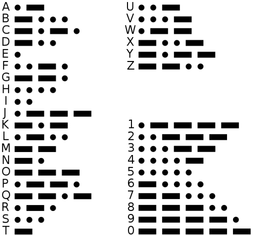
Morsejeva abeceda

Morsov kod

- $\Sigma = \{A, B, \ldots, Z\} \cup \{0, 1, \ldots, 9\} \cup \{\text{␣}\}$
- $\Sigma_c = \{\cdot, -, \square\}$
- $\kappa_S : \Sigma \to \Sigma_c^*$

Kodna funkcija črk κ_s

- Vrednosti funkcije so določene s tabelo



- Dodatno $\kappa_s(\textvisiblespace) = \square\square\square$

Morsejeva abeceda

Morsov kod

- $\Sigma = \{A, B, \dots, Z\} \cup \{0, 1, \dots, 9\} \cup \{␣\}$
- $\Sigma_c = \{\cdot, -, \square\}$
- $\kappa_s : \Sigma \to \Sigma_c^*$
- $\kappa(w) = \kappa_s(a_1)\square\square\square\square\kappa_s(a_2)\square\square\square\square \cdots \kappa_s(a_n)$

Morsejeva abeceda

Morsov kod

- $\Sigma = \{A, B, \ldots, Z\} \cup \{0, 1, \ldots, 9\} \cup \{\textvisiblespace\}$
- $\Sigma_c = \{\cdot, -, \square\}$
- $\kappa_s : \Sigma \to \Sigma_c^*$
- $\kappa(w) = \kappa_s(a_1)\square\square\square\square\kappa_s(a_2)\square\square\square\square \cdots \kappa_s(a_n)$

Primer kodiranja z Morsejevo abecedo

$$
\begin{aligned}
\kappa(\mathrm{SOS}) &= \kappa_s(S)\square\square\square\square\kappa_s(O)\square\square\square\square\kappa_s(S) \\
&= \cdot\square\cdot\square\cdot\square\square\square - \square - \square - \square\square\square\cdot\square\cdot\square\cdot
\end{aligned}
$$

Stiskanje podatkov

Definicija

Stiskanje je kodiranje K za katerega velja

$$\exists n \in \mathbb{N}\, \forall w \in \Sigma^* \colon |w| \geq n \implies |\kappa(w)| \ll |w| \,.$$

Stiskanje podatkov

Primer stiskanja niza

- Za abecedo vzemimo $\Sigma = \{a, b, c\}$ in stisnimo niz

$$w = cababcccababcccab$$

Stiskanje podatkov

Primer stiskanja niza

- Za abecedo vzemimo $\Sigma = \{a, b, c\}$ in stisnimo niz

$$w = cababcccababcccab$$

- Uvedemo oznaki $A = ab$ in $B = ccc$

$$w = cAABAABA$$

Stiskanje podatkov

Primer stiskanja niza

- Za abecedo vzemimo $\Sigma = \{a, b, c\}$ in stisnimo niz

$$w = cababcccababcccab$$

- Uvedemo oznaki $A = ab$ in $B = ccc$

$$w = cAABAABA$$

- Uvedemo novo spremeljivko $C = AAB$

$$w = cCCA$$

Stiskanje podatkov

Primer stiskanja niza

Prešnji postopek zapišemo na sledeč način

$$S \to cCCA,$$
$$A \to ab,$$
$$B \to ccc,$$
$$C \to AAB.$$

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 z 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 z kontekstno-neodvisno gramatiko

Stiskanje niza $w = cababcccababcccab$

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

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

Velja, da je

$$L(G) = \{w\}$$