

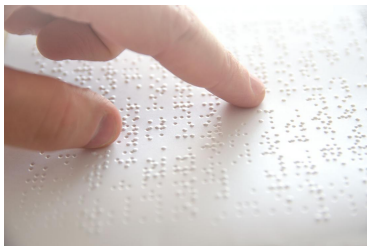
# Kontekstno-neodvisne gramatike za kodiranje in stiskanje podatkov

Janez Podlogar

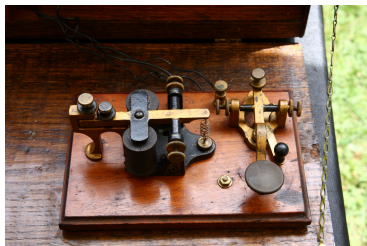
Univerza v Ljubljani, Fakulteta za matematiko in fiziko

21. 11. 2022

# Kodiranje in kod



**Slika:** Braillova pisava



**Slika:** Telegrafska tipka in zvočnik

# 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 z zaporedjem signalov:

# Morsejeva abeceda

*Morsejeva abeceda* je kodiranje črk, števil in ločil z zaporedjem signalov:

- Dolžina kratkega signala je ena enota

# Morsejeva abeceda

*Morsejeva abeceda* je kodiranje črk, števil in ločil z zaporedjem 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 z zaporedjem 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 z zaporedjem 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 oz. en dolgi signal

# Morsejeva abeceda

*Morsejeva abeceda* je kodiranje črk, števil in ločil z zaporedjem 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 oz. en dolgi signal
- Presledek med besedami je tišina dolga sedem kratkih signalov

# Abeceda in nizi na abecedi

## Definicija

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

# Abeceda in nizi na abecedi

## Definicija

- *Abeceda* je končna neprazna množica  $\Sigma$
- *Množica vseh končnih nizov abecede*  $\Sigma$  označimo s  $\Sigma^*$
- *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

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

# Kodiranje in dekodiranje

## Definicija

- *Kodiranje nizov abecede  $\Sigma$*  je injektivna funkcija  $\kappa: \Sigma^* \rightarrow \Sigma_c^*$

# Kodiranje in dekodiranje

## Definicija

- *Kodiranje nizov abecede*  $\Sigma$  je injektivna funkcija  $\kappa: \Sigma^* \rightarrow \Sigma_c^*$
- *Dokodiranje kodiranja*  $\kappa$  je funkcija  $\delta: C \subseteq \Sigma_c^* \rightarrow \Sigma^*$ , da velja

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

# Morsejeva abeceda

Morsov kod

# Morsejeva abeceda

## Morsov kod

- $\Sigma = \{A, B, \dots, Z\} \cup \{0, 1, \dots, 9\} \cup \{\_ \}$



# Morsejeva abeceda

## Morsov kod

- $\Sigma = \{A, B, \dots, Z\} \cup \{0, 1, \dots, 9\} \cup \{\_ \}$
- $\Sigma_c = \{., -, \square\}$

# Morsejeva abeceda

## Morsov kod

- $\Sigma = \{A, B, \dots, Z\} \cup \{0, 1, \dots, 9\} \cup \{\_ \}$
- $\Sigma_c = \{., -, \square\}$
- $\kappa_S: \Sigma \rightarrow \Sigma_c^*$

## Kodna funkcija črk $\kappa_s$

- Vrednosti funkcije so določene s tabelo

A	•	—					
B	—	—	•	•			
C	—	—	•		•		
D	—	—	•			•	
E	•						
F	•	—	—	•			
G	•	—	—	•			
H	•	•					
I	•						
J	•	—	—	—			
K	—	•	•				
L	—	•	•				
M	—	•					
N	—	•					
O	—	—	•	•			
P	—	—	•	•			
Q	•	•	•	•			
R	•	•	•	•			
S	•	•	•	•			
T	—	—	—	—			
U	•	•	•				
V	•	•	•				
W	•	•	•				
X	•	•	•	•			
Y	•	•	•	•			
Z	•	•	•	•			
1	•	—	—	—	—		
2	•	—	—	—	—		
3	•	—	—	—	—		
4	•	—	—	—	—		
5	•	—	—	—	—		
6	•	—	—	—	—		
7	•	—	—	—	—		
8	•	—	—	—	—		
9	•	—	—	—	—		
0	•	—	—	—	—		

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

# Morsejeva abeceda

## Morsov kod

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

# Morsejeva abeceda

## Morsov kod

- $\Sigma = \{A, B, \dots, Z\} \cup \{0, 1, \dots, 9\} \cup \{\_ \}$
- $\Sigma_c = \{., -, \square\}$
- $\kappa_S: \Sigma \rightarrow \Sigma_c^*$
- $\kappa(w) = \kappa_S(a_1)\square\square\square\kappa_S(a_2)\square\square\square\cdots\kappa_S(a_n)$

## Primer kodiranja z Morsejevo abecedo

$$\begin{aligned}\kappa(\text{SOS}) &= \kappa_S(\text{S})\square\square\square\kappa_S(\text{O})\square\square\square\kappa_S(\text{S}) \\ &= .\square.\square.\square\square\square - \square - \square - \square\square\square.\square.\square.\end{aligned}$$

# Stiskanje podatkov

## Definicija

*Stiskanje* je kodiranje  $K$  za katerega velja

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

# 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 oznako  $C = AAB$

$$w = cCCA$$

# Stiskanje podatkov

## Primer stiskanja niza

Prešnji postopek zapišemo na sledeč način

$$S \rightarrow cCCA,$$

$$A \rightarrow ab,$$

$$B \rightarrow ccc,$$

$$C \rightarrow AAB$$

# Kontekstno-neodvisne gramatike

## Definicija

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

# Kontekstno-neodvisne gramatike

## Definicija

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

- $V$  končna množica *nekončnih simbolov*

# Kontekstno-neodvisne gramatike

## Definicija

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

- $V$  končna množica *nekončnih simbolov*
- abeceda  $\Sigma$  množica *končnih simbolov*

# Kontekstno-neodvisne gramatike

## Definicija

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

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

# Kontekstno-neodvisne gramatike

## Definicija

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

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

# Kontekstno-neodvisne gramatike

## Definicija

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

- $V$  končna množica *nekončnih simbolov*
- abeceda  $\Sigma$  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

# Stiskanja niza z kontekstno-neodvisno gramatiko

Stiskanje niza  $w = cababcccababcccab$

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

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

# Stiskanja niza z kontekstno-neodvisno gramatiko

Stiskanje niza  $w = cababcccababcccab$

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

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

# Stiskanja niza z kontekstno-neodvisno gramatiko

Stiskanje niza  $w = cababcccababcccab$

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

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

# Stiskanja niza z kontekstno-neodvisno gramatiko

Stiskanje niza  $w = cababcccababcccab$

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

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

# Stiskanja niza z kontekstno-neodvisno gramatiko

Stiskanje niza  $w = cababcccababcccab$

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

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

Velja, da je

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