

KONTEKSTNO-NEODVISNE GRAMATIKE ZA KODIRANJE IN STISKANJE PODATKOV

JANEZ PODLOGAR

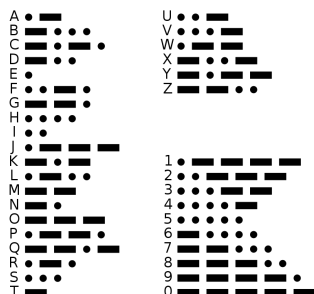
POVZETEK. V delu predstavimo motivacijo in osnovne pojme, ki so potrebni za obravnavanje stiskanja podatkov s kontekstno-neodvisnimi gramatikami.

1. KODIRANJE PODATKOV

Zapis informacije v neki obliki ni primeren za vsakršno rabo. Besedilo, zapisano z pismenkami, je neberljivo za slepe osebe, saj je komunikacijski kanal v tem primeru vid. Prav tako pisanega besedila v pravotni obliki ni moč poslati z telegrafom. V tem primeru je komunikacijski kanal žica in pismenke se po njej ne morejo sprehoditi. V obeh primerih je informacija, ki bi jo radi prenesli v neprimerni obliki. V prvem primeru je potrebno besedilo zapisati z Braillovo pisavo. V drugem primeru pa je potrebno besedilo pretvoriti v električni signal. Spreminjanje zapisa sporočila pravimo *kodiranje*, sistemu pravil, po katerem se kodiranje opravi, pa *kod*.

Primer 1.1. *Morsejeva abeceda* je kodiranje črk, števil in ločil s pomočjo zaporedja kratkih in dolgih signalov:

- Dolžina kratkega signala je ena enota.
- Dolgi signal je trikrat daljši od kratkega signala.
- Razmiki med signali znotraj znaka so dolžine kratkega signala.
- Presledki med znaki so dolgi tri kratke signale oz. en dolgi signal.
- Presledki med besedami so dolgi sedem kratkih signalov.



SLIKA 1. Mednarodna Morsejeva abeceda

Prvotni namen Morsejeve abecede je komunikacija preko telegrama, saj komunikacijski kanal dovoljuje le električne signale in tišino med njimi. Kodiranje črk je takšno, da imajo črke z višjo frekvenco (v angleškem jeziku) krajši zapis, s tem se dolžina kodiranega sporočila skrajša in posledično tudi čas prenosa.

◇

Definicija 1.2. *Abeceda* je vsaka neprazna množica Σ . *Množica vseh besed na abecedi* Σ je:

$$\Sigma^* = \{ a_1 a_2 a_3 \cdots \mid \forall i : a_i \in \Sigma \} \cup \{ \varepsilon \},$$

kjer ε prazna beseda.

V splošnem ima lahko abeceda poljubno kardinalnost, v diplomski nalogi se bomo srečali le z končnimi abecedami.

Primer 1.3. Naj bo $\Sigma = \{a, b, c\}$ abeceda, potem je:

$$ab \in \Sigma^*$$

$$ccc \in \Sigma^*$$

$$cababcccababcccab \in \Sigma^*$$

◇

2. KOMPRESIJA PODATKOV

3. KONTEKSTNO-NEODVISNE GRAMATIKE