Azbuka i kodovi

- *Konačna azbuka V* – konačan neprazni skup proizvoljnih simbola
- *slova* ili *simboli* – elementi skupa V
- *reči* – konačne niske slova iz V.
- *Prazna reč*, u oznaci λ – reč koja ne sadrži slova.
- V^* – skup svih reči nad V
- $\lambda \in V^*$, za svako V.
  $V^+ = V^* \setminus \{\lambda\}$
- *jezik L* – proizvoljan skup reči iz V^*.
- *dužina reči* $|P|$– broj slova u reči P. Važi $|\lambda| = 0$.
- $P^i, i > 0$ označava i puta dopisanu reč P. Po definiciji, $P^0 = \lambda$.
- Broj reči dužine d u azbuci koja ima n znakova je n^d.

Primeri azbuka i jezika nad njima

- Za azbuku V_1={a, b} možemo da definišemo sledeće jezike:
  - L_1={a, b, λ} – jezik sa rečima dužine ≤ 1
  - L_2={$a^i b^i$ |i $\in${0,1}} – jezik sa rečima dužine ≤ 2
  - ...
- Za azbuku V_2={0, 1, 2, 3, 4, 5, 6, 7, 8, 9 } reči nad V_2^+ su brojevi u dekadnom brojčanom sistemu (sa eventualnim vodećim nulama)
- Za azbuku V_3={0, 1 } reči nad V_3^+ su brojevi u binarnom brojčanom sistemu (sa eventualnim vodećim nulama)

Kodovi

Neka su date azbuke

V_1 = { a_1, a_2, ..., a_m}, m≥1,

V_2 = { b_1, b_2, ..., b_n}, n≥2

i jezici

$L_1 \subset V_1^*$, L_1 = { p_1, p_2, ..., p_r }, r≥1, i

$L_2 \subset V_2^*$, L_2 = { q_1, q_2, ..., q_s }, s≥1.

Tada možemo da uvedemo sledeće definicije:

1. **Funkcija kodiranja** je svaka funkcija f definisana sa $f : L_1 \longrightarrow L_2$.
2. **Funkcija dekodiranja** je funkcija $g : L_2 \longrightarrow L_1$, $g = f^{-1}$, ako f^{-1} postoji.
3. *Kod* jezika L_1 u azbuci V_2 je skup vrednosti K⊆L_2,$K = \{q_j \mid \exists i : f(p_i) = q_j\}$. *Kodiranje* predstavlja izračunavanje vrednosti $f(p_i)$. *Dekodiranje* predstavlja izračunavanje vrednosti $g(q_j)$.
4. Kod je *jednoznačan* akko je funkcija f 1-1. U suprotnom je kod *višeznačan*.
5. Broj simbola u azbuci V_2 se naziva *osnova koda*. Reč q_i iz jezika L_2 se naziva *kodna reč*.
6. Kod je *potpun* kada obuhvata sve reči odredjene dužine u jeziku L_2.
7. Kod je *ravnomeran* ako $\exists l\ \forall j : \mid q_j \mid = l$, tj. ako je dužina svih kodnih reči u jeziku L_2 ista. U suprotnom, kod je *neravnomeran*. Da bi kod dužine d reči P bio ravnomeran mora da važi |P|≤n^d gde je n broj simbola u azbuci V_2. Ako važi |P|= n^d tada je u pitanju *potpun ravnomeran kod*. Dužina reči ravnomernog koda se naziva *mesnost* koda.

Primeri:

1. Neka azbuka $V_1 = \{ +, -, *, / \}$ sadrži oznake elementarnih aritmetičkih operacija. Za kodiranje odgovarajućeg jezika $L_1 = \{ +, -, *, / \}$ u binarnoj azbuci dovoljno je uzeti reči dužine 2. Neka moguća kodiranja su prikazana u narednoj tabeli. Svi prikazani kodovi su ravnomerni i potpuni.

Reč u jeziku	Kod 1	Kod 2	Kod 3	Kod 4	Kod 5	Kod 6
$+$	00	00	00	00	00	00
-	01	01	10	10	11	11
$*$	10	11	01	11	01	10
/	11	10	11	01	10	01

2. Neka azbuka V_1 sadrži sledeće simbole: $V_1 = \{ \triangleleft, \triangleright, \oplus, \ominus, \otimes, \oslash \}$. Za kodiranje odgovarajućeg jezika $L_1 = \{ \triangleleft, \triangleright, \oplus, \ominus, \otimes, \oslash \}$ u binarnoj azbuci dovoljno je uzeti reči dužine 3. Jedno moguće kodiranje je prikazano u narednoj tabeli. Dobijeni kod, bez obzira na izbor funkcije, nije potpun.

Reč u jeziku	Kod
$\triangleleft$	000
$\triangleright$	001
$\oplus$	010
$\ominus$	011
$\otimes$	100
$\oslash$	101

3. Ukoliko želimo da zapišemo reči živog jezika tada za azbuku V_1 moramo da uzmemo sve znake pisma (velika i mala slova, cifre, interpunkcijske i specijalne znake). Dužina kodnih reči će zavisiti od broja simbola u azbuci, odnosno broja reči u odgovarajućem (formalnom) jeziku L_1. U ranijem periodu razvoja računarstva postojalo je više različitih kodova, ali se danas koriste kodovi sa dužinom reči 7, 8 ili 16.