Matematičke osnove čuvanja podataka

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.
  $V^+ = V^* \backslash \{\lambda\}$
- *jezik* L – proizvoljan skup reči nad V^*.
- *dužina reči* $|\mathrm{P}|$– broj slova u reči P. Važi $|\lambda| = 0$.
- $P^i, i > 0$ označava i puta dopisanu reč P. Po definiciji, $\mathrm{P}^0 = \lambda$.
- 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=0,1} – jezik sa rečima dužine ≤ 2
  - ...
- Za azbuku V_2={0, 1, 2, 3, 4, 5, 6, 7, 8, 9 } reči nad V^+ su brojevi u dekadnom brojnom sistemu (sa eventualnim vodećim nulama)
- Za azbuku V_3={0, 1 } reči nad V^+ su brojevi u binarnom brojnom sistemu (sa eventualnim vodećim nulama)

Kodovi

Neka su date azbuke

$V_1 = \{ a_1, a_2, ..., a_m\}, m \geq 1,$

$V_2 = \{ b_1, b_2, ..., b_n\}, n \geq 2$

i jezici

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

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

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 $K \subseteq L_2, K = \{q_j \mid \exists i : f(p_i) = q_j\}$. *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 $|P| \leq n^d$ 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.

Brojčani sistemi

1. nepozicioni – znak koji označava cifru ima istu vrednost bez obzira na poziciju u zapisu broja
2. pozicioni – vrednost znaka koji predstavlja cifru zavisi i od izgleda znaka i od pozicije cifre u zapisu broja.

Broj različitih cifara pozicionog brojčanog sistema se naziva *osnova brojčanog sistema*.

Neka skup S sadrži $N > 1$ cifara. Brojčana vrednost X u pozicionom sistemu sa osnovom N piše se u obliku niske cifara, uz poštovanje sledećih pravila:

1. Brojčana vrednost X u sistemu sa osnovom N

   $$(X)_N = \sum_{i=-m}^{n} V(x_i)$$

   gde su x_i cifre brojčanog sistema, $V(x_i)$ vrednost cifre x_i u zapisanoj nisci cifara, a i mesto cifre u zapisanoj nisci cifara ($i \in [-m, n]$).

2. $V(x_i) = x_i \cdot N^i$. Odavde

   $$(X)_N = \sum_{i=-m}^{n} x_i \cdot N^i = x_n N^n + \ldots + x_0 N^0 + x_{-1} N^{-1} + \ldots + x_{-m} N^{-m}$$

   Sve operacije u ovom izrazu se vrše u brojčanom sistemu sa osnovom N

3. Po konvenciji ne pišu se osnova i stepen, a celobrojni i razlomljeni deo se razdvajaju zarezom

   $$(X)_N \equiv x_n x_{n-1} \ldots x_0 , x_{-1} \ldots x_{-m}$$

Primeri:

1. Brojčani sistem kod koga je $N=10$, $S=\{0,\ldots,9\}$ se naziva *dekadni sistem*.
2. Brojčani sistem kod koga je $N=8$, $S=\{0,1,2,3,4,5,6,7\}$ se naziva *oktalni sistem*. Primeri brojeva zapisanih u ovom sistemu su 123,456 i 243. Vrednost ovih brojeva u dekadnom sistemu je:
   - $(243)_8 = V(2_2)+V(4_1)+V(3_0) = 2\times8^2 + 4\times8^1 + 3\times8^0 = (128)_{10}+(32)_{10}+(3)_{10} = (163)_{10}$
   - $(123{,}456)_8 = V(1_2)+V(2_1)+V(3_0)+V(4_{-1})+V(5_{-2})+V(6_{-3}) = 1\times8^2 + 2\times8^1 + 3\times8^0 + 4\times8^{-1} + 5\times8^{-2} + 6\times8^{-3} = (83{,}589843...)_{10}$
3. Brojčani sistem kod koga je $N=2$, $S=\{0,1\}$ se naziva *binarni sistem*. Primer broja u binarnom sistemu je 1011110. Njegova vrednost je:
   - u dekadnom sistemu:
     $(1011110)_2 = 1\times2^6+0\times2^5+1\times2^4+1\times2^3+1\times2^2+1\times2^1+0\times2^0$
     $= (64)_{10}+(16)_{10}+(8)_{10}+(4)_{10}+(2)_{10} = (94)_{10}$
   - u oktalnom sistemu:
     $(1011110)_2 = 1\times2^6+0\times2^5+1\times2^4+1\times2^3+1\times2^2+1\times2^1+0\times2^0$
     $= (100)_8+(20)_8+(10)_8+(4)_8+(2)_8 = (136)_8$
4. Brojčani sistem kod koga je $N=16$, `S={0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F}` se naziva *heksadekadni sistem*. Primer broja u heksadekadnom sistemu je CDE92. Njegova vrednost u dekadnom sistemu je

   $(CDE92)_{16} = 12\times16^4+13\times16^3+14\times16^2+9\times16^1+2\times16^0 =$
   $(786432)_{10}+(53248)_{10}+(3584)_{10}+(144)_{10}+(2)_{10} = 843410_{10}$

Prevodjenje brojeva

$(X)_N \equiv x_n x_{n-1} \ldots x_0, x_{-1} \ldots x_{-m}$ – zapisan u sistemu sa osnovom N.

$(X)_M \equiv y_p y_{p-1} \ldots y_1 y_0, y_{-1} \ldots y_{-q}$ – zapisan u sistemu sa osnovom M.

prevodjenje broja X iz sistema sa osnovom N u sistem sa osnovom M – postupak odredjivanja vrednosti broja X u sistemu sa osnovom M.

Mešoviti brojevi se prevode tako što se odvojeno prevode celi i razlomljeni delovi i tako dobijeni prevodi spoje.

Prevodjenje celih brojeva

Dato X u sistemu sa osnovom N

$$(X)_N \equiv x_n x_{n-1} \ldots x_1 x_0 = \sum_{i=0}^{n} x_i \, N^i$$

Ista vrednost u sistemu sa osnovom M

$$(X)_M \equiv y_p \ldots y_1 y_0 = \sum_{i=0}^{p} y_i \, M^i$$

Važi $(X)_N = (X)_M$

$$\frac{(X)_N}{M} = \frac{y_0}{M} + \sum_{i=1}^{p} y_i \, M^{i-1}$$

Deljenjem broja X sa osnovom M dobija celobrojni deo količnika

$$X_1 = \sum_{i=1}^{p} y_i \, M^{i-1}$$

i ostatak deljenja y_0 koji predstavlja cifru na jediničnom mestu u broju X predstavljenom u sistemu sa osnovom M.

Rekurentna formula

$$\begin{aligned} X_i/M &= X_{i+1} + y_i/M \\ X_0 &= X \end{aligned}$$

pri čemu se aritmetičke operacije izvode u sistemu sa osnovom N.

Šematski postupak

i	0	1	2	...	p
X_i	X_0	X_1	X_2	...	X_p
y_i	y_0	y_1	y_2	...	y_p

$\longleftarrow$ smer čitanja cifara

X_{i+1} – celobrojni deo količnika X_i/M
y_i – ostatak pri ovom deljenju
Postupak se ponavlja sve dok se ne dodje do broja $X_{p+1} = 0$ zdesna ulevo

Primeri:

1. $94_{10} \longrightarrow (1011110)_2$

i	0	1	2	3	4	5	6
X_i	94	47	23	11	5	2	1
y_i	0	1	1	1	1	0	1

2. $AB_{16} \longrightarrow (10101011)_2$

i	0	1	2	3	4	5	6	7
X_i	AB	55	2A	15	A	5	2	1
y_i	1	1	0	1	0	1	0	1

3. $(101110)_2 \longrightarrow (2E)_{16}$

i	0	1	2
X_i	101110	10	0
y_i	1110	10	0

Prevodjenje razlomljenog dela

$$(X)_N \equiv 0, x_{-1} x_{-2} \ldots x_{-m} = \sum_{i=1}^{m} x_{-i}\; N^{-1}$$

Isti broj X u sistemu sa osnovom M se zapisuje u obliku

$$(X)_M \equiv 0, y_{-1} y_{-2} \ldots = y_{-1}\; M^{-1} + y_{-2}\; M^{-2} + \ldots$$

Izjednačavanjem ovih jednakosti $(X)_N = (X)_M$ i množenjem sa osnovom M

$$(X)_N \times M = y_{-1} + y_{-2}\; M^{-1} + y_{-3}\; M^{-2} + \ldots$$

dobija se zbir cifre y_{-1} i razlomljenog dela

$$X_{-1} = y_{-2}\; M^{-1} + y_{-3}\; M^{-2} + \ldots$$

Množenjem ove jednakosti sa M dobija se cifra y_{-2} itd.

Rekurentna formula

$$\begin{aligned} X_i * M &= y_{-(i+1)} + X_{-(i+1)} \\ X_{-0} &= X \end{aligned}$$

Aritmetičke operacije se izvode u sistemu sa osnovom N

Šematski postupak

i	0	1	2	…	q
X_{-i}	X_{-0}	X_{-1}	X_{-2}	…	X_{-q}
y_{-i}	0	y_{-1}	y_{-2}	…	y_{-q}

smer čitanja cifara $\longrightarrow$

Primeri:

1. $(0{,}84375)_{10} \longrightarrow (0{,}11011)_2$

i	0	1	2	3	4	5
X_{-i}	0,84375	0,68750	0,3750	0,750	0,50	0,00
y_{-i}	0	1	1	0	1	1

2. $(0,4)_{10} \approx (0,011001100\ldots)_2$

i	0	1	2	3	4	5	6	7	8	9
X_{-i}	0,4	0,8	0,6	0,2	0,4	0,8	0,6	0,2	0,4	0,8
y_{-i}	0	0	1	1	0	0	1	1	0	0

3. $(0,4)_{10} \approx (0,1212\ldots)_4$

i	0	1	2	3	4
X_{-i}	0,4	0,6	0,4	0,6	0,4
y_{-i}	0	1	2	1	2

Specijalni slučaj kodiranja

Ako važi $N = M^s, s > 1$, pri prevodjenju brojeva izmedju sistema sa osnovama N i M se koristi tvrdjenje

Vrednost broja X u sistemu sa osnovom N zapisana u sistemu sa osnovom M je identična zapisu koji se dobija kodiranjem cifara broja X u sistemu sa osnovom M. Prevodjenje mešovitih brojeva se vrši tako što se posebno prevedu celobrojni i razlomljeni deo i od dobijenih prevoda formira željeni prevod.

Primeri:

1. $(54,12)_8 \longrightarrow (....,....)_2$.

   Binarni zapisi oktalnih cifara

Oktalna cifra	Binarni zapis
0	000
1	001
2	010
3	011
4	100
5	101
6	110
7	111

   Prema tvrdjenju važi $(54,12)_8 \longrightarrow (101100,001010)_2$.

   Sa druge strane $(54,12)_8 = 5\times8^1 + 4\times8^0 + 1\times8^{-1} + 2\times8^{-2} =$
   $(1\times2^2 + 0\times2^1 + 1\times2^0)\times2^3 + (1\times2^2 + 0\times2^1 + 0\times2^0)\times2^0 +$
   $(1\times2^0)\times2^{-3} + (1\times2^1 + 0\times2^0)\times2^{-6} = 1\times2^5 + 1\times2^3 + 1\times2^2 + 1\times2^{-3} + 1\times2^{-5} =$
   $= (101100,001010)_2$

2. $(\mathrm{ABC,DE})_{16} \longrightarrow (....,....)_8$

   (a) ABC,DE se prevodi u binarni sistem

   $$\begin{array}{lcl} (\mathrm{ABC,DE})_{16} & \longrightarrow & (1010|1011|1100,1101|1110)_2 \\ & \longrightarrow & (101010111100,11011110)_2 \end{array}$$

   (b) Dobijeni binarni broj se prevede i oktalni sistem.

   $$\begin{array}{lcl} (101010111100,11011110)_2 & \longrightarrow & \\ (101|010|111|100,110|111|100)_2 & \longrightarrow & (5274,674)_8 \end{array}$$

Zadaci:

Prevesti sledeće brojeve:

1. $(21012)_3 \longrightarrow (.....)_{16}$
2. $(201)_3 \longrightarrow (.....)_2$
3. $(634)_7 \longrightarrow (.....)_{16}$, bez medjuprevodjenja u dekadni sistem
4. $(0{,}25)_{10} \longrightarrow (.....)_{16}$.
5. $(0{,}66)_{10} \longrightarrow (.....)_6$.
6. $(14{,}34)_{10} \longrightarrow (.....)_5$.
7. $(AB7F)_{16} \longrightarrow (.....)_4$.
8. $(3220)_4 \longrightarrow (.....)_8$.
9. $(0{,}3DC)_{16} \longrightarrow (.....)_8$.
10. $(3FCED0{,}179A)_{16} \longrightarrow (.....)_4$.