

4. Digitalizácia

Uchovanie informácie súvisí s rozvojom komunikácie medzi ľuďmi. Aby sa informácie mohli šíriť aj na väčšie vzdialenosti, bolo treba nájsť iný spôsob, ako ústne podanie a uchovanie v pamäti človeka.

Myšlienka zapisovania a uchovávanía písaných záznamov súvisí so vznikom symbolov potrebných na zápis - vznikom písma.

S rozvojom fyziky sa na uchovávanie informácií začal využívať celý rad fyzikálnych zákonitostí:

- elektro-mechanický záznam (dierne štítky, dierne pásky, vinylové platne),
- magnetické pole a magnetizácia látok (pevné disky, diskety),
- optika - vlastnosti monochromatického svetla (lasera) v zápise na kompaktné disky, DVD disky, Blu-ray disky, atď.

Údaje sú v technických zariadeniach uchovávané ako postupnosti **núl** a **jednotiek**. Takýto spôsob zápisu, ktorý používa na zápis informácií nuly a jednotky sa nazýva **binárny kód**.

Kódovanie informácií

Cieľom kódovania je, aby sme mohli prenášať informáciu. Často sa musíme prispôbiť možnostiam technického zariadenia (preto vznikla napr. Morseova abeceda ako jazyk pre telegraf), alebo možnostiam ľudí zapojených do komunikácie (preto vzniklo Braillovo písmo pre nevidiacich).

Kódovanie informácií je ľubovoľná vopred dohodnutá a všeobecne známa množina pravidiel, ktorá dovoľuje informáciu vyjadriť tak, aby sa dala uchovať, alebo šíriť.

Šifrovanie sa používa všade tam, kde treba utajiť obsah komunikácie. Existuje veľmi veľa metód na tajné šifrovanie (a metód na dešifrovanie).

Binárny kód.

Ľudia medzi sebou používajú na komunikáciu reč, ktorá pozostáva zo slov tvorených jednotlivými písmenami. Počítače komunikujú – prenášajú informácie - v číslach. Informácie, ktoré v nich človek uchováva, musia byť prekódované do im zrozumiteľného jazyka.

Pamäť počítača si môžeme predstaviť ako milióny miniatúrnych prepínačov, z ktorých každý je buď vypnutý (pre znak 0) alebo zapnutý (pre znak 1). Dlhé postupnosti prepínačov predstavujú rôzne informácie.

Počítače preto používajú zvláštny spôsob kódovania informácií – **binárny kód**. *Sú to postupnosti dvoch znakov – 0 a 1.*

Informácie zapísané v binárnom kóde nazývame **digitálne informácie**.

Digitalizácia - prevod informácie z reálneho sveta (analogového) do binárneho kódu podľa dohodnutých pravidiel.

5. Reprezentácia textovej informácie

Keď chceme pracovať s textovou informáciou v počítači musíme ju najprv nejakým spôsobom vedieť reprezentovať. Veľmi výhodnou možnosťou je **rozložiť textovú informáciu na jednotlivé znaky**.

Znakom nie sú len malé a veľké písmená, ale aj číslice, špeciálne znaky – čiarka, bodka, plus, pomlčka, dvojbodka, matematické operácie plus, mínus aj takzvané neviditeľné znaky. To sú napríklad medzera, tabulátor, koniec riadku, koniec súboru a iné.

Keď máme text rozložený na jednotlivé znaky, zistíme, že týchto znakov nie je nejako veľa. Sú len rôzne usporiadané a tým vytvárajú text. Teraz každému znaku priradíme nejaký kód a následne v súbore celý text zapíšeme **ako postupnosť kódov znakov textu**.

Takéto priradenie kódov znakom nazývame **kódová tabuľka**. V minulosti boli znaky kódované pomocou 7 bitov. To umožnilo kódovať 128 znakov. Takéto kódovanie označujeme **ASCII (American Standard Code for Information Interchange)**.

Toto kódovanie celkom stačilo pre anglickú abecedu a množinu základných symbolov. Ale iné jazyky potrebujú písať text s diakritikou a preto sa začalo používať na kód znaku **8 bitov = 1 byte**. Teda dalo sa zakódovať 256 rôznych znakov. Prvých 128 znakov bolo spoločných, zodpovedali pôvodnému ASCII kódovaniu. Zvyšných 128 znakov si každé kódovanie vybralo podľa svojich potrieb. Vznikli **rozšírené ASCII tabuľky**.

Spočiatku bol takýto systém výhodný. Umožňoval písať texty v národných abecedách s diakritikou. Ale neskôr s rozvojom počítačových sietí dochádzalo čoraz častejšie k problémom. Každé kódovanie malo znaky s kódmi 129 až 255 rôzne. To spôsobilo, že ak sme napísali text v programe s nastaveným kódovaním napríklad pre strednú Európu a poslali sme ho niekomu, kto používal kódovanie pre západnú Európu, tak znaky anglickej abecedy sa zobrazili korektne, ale znaky s diakritikou a rôzne špeciálne znaky sa mohli zobrazíť úplne nezmyselne. Pre Strednú Európu vzniklo viacero kódovaní: windows-1250, ISO-8859-2, MacCE, IBM-852 a pod.

Dec	Hex	Znak	Význam
0	00	NUL	
1	01	SOH	Start of Header
2	02	STX	Start of Text
3	03	ETX	End of Text
4	04	EOT	End of Transmission
5	05	ENQ	Enquiry
6	06	ACK	Acknowledge
7	07	BEL	Bell
8	08	BS	Backspace
9	09	HT	Horizontal Tab
10	0A	LF	Line Feed
11	0B	VT	Vertical Tab
12	0C	FF	Form Feed
13	0D	CR	Carriage Return
14	0E	SO	Shift Out
15	0F	SI	Shift In
16	10	DLE	Data Link Escape
17	11	DC1	(XOn)
18	12	DC2	(XOff)
19	13	DC3	
20	14	DC4	
21	15	NAK	Negative Acknowledge
22	16	SYN	Synchronous Idle
23	17	ETB	End of Transmission Block
24	18	CAN	Cancel
25	19	EM	End of Medium
26	1A	SUB	Substitute
27	1B	ESC	Escape
28	1C	FS	File Separator
29	1D	GS	Group Separator
30	1E	RS	Record Separator
31	1F	US	Unit Separator

Dec	Hex	Znak
32	20	SPC
33	21	!
34	22	"
35	23	#
36	24	\$
37	25	%
38	26	&
39	27	'
40	28	(
41	29)
42	2A	*
43	2B	+
44	2C	,
45	2D	-
46	2E	.
47	2F	/
48	30	0
49	31	1
50	32	2
51	33	3
52	34	4
53	35	5
54	36	6
55	37	7
56	38	8
57	39	9
58	3A	:
59	3B	;
60	3C	<
61	3D	=
62	3E	>
63	3F	?

Dec	Hex	Znak
64	40	@
65	41	A
66	42	B
67	43	C
68	44	D
69	45	E
70	46	F
71	47	G
72	48	H
73	49	I
74	4A	J
75	4B	K
76	4C	L
77	4D	M
78	4E	N
79	4F	O
80	50	P
81	51	Q
82	52	R
83	53	S
84	54	T
85	55	U
86	56	V
87	57	W
88	58	X
89	59	Y
90	5A	Z
91	5B	[
92	5C	\
93	5D]
94	5E	^
95	5F	_

Dec	Hex	Znak
96	60	`
97	61	a
98	62	b
99	63	c
100	64	d
101	65	e
102	66	f
103	67	g
104	68	h
105	69	i
106	6A	j
107	6B	k
108	6C	l
109	6D	m
110	6E	n
111	6F	o
112	70	p
113	71	q
114	72	r
115	73	s
116	74	t
117	75	u
118	76	v
119	77	w
120	78	x
121	79	y
122	7A	z
123	7B	{
124	7C	
125	7D	}
126	7E	~
127	7F	DEL

Aby sa predišlo podobným problémom bolo vytvorené kódovanie s použitím 16 bitov s názvom **UNICODE**. V tomto kódovaní je možné zakódovať cez 65 536 rôznych znakov, čo umožňuje zakódovať znaky všetkých abecied pomocou jednej medzinárodnej tabuľky. Toto kódovanie zabezpečuje, že ten istý znak má rovnaký kód v každej krajine a každom type počítača.

Nevýhodou tohto kódovania je, že znaky, ktoré sme predtým vedeli zakódovať iba ôsmymi bitmi. V kódovaní Unicode sú kódované 16 bitmi, a teda zaberajú viac pamäte ako kód ASCII.

Istým vylepšením tohto kódovania je kódovanie **UTF-8** (8-bitový Unicode Transformation Format). V tomto kódovaní je prvých 128 znakov tabuľky ASCII, zakódovaných pomocou 8 bitov a zvyšné znaky sú zakódované 16, 24, 32, 40 až 48 bitmi.

Digitalizácia informácie

Zjednodušene môžeme povedať, že svet okolo nás je plný analógových informácií. Keď ich chceme spracovať v počítači, musíme ich previesť na digitálne. Tento proces sa nazýva **digitalizácia**. Pôvodná analógová informácia musí byť rekonštruovateľná z výslednej digitálnej informácie. Z povahy digitalizácie vyplýva, že určité podrobnosti informácie sa pri digitalizácii stratia.

Základom digitalizácie (prevod spojitého signálu na diskrétny) sú tri procesy: **vzorkovanie**, **kvantovanie**, **kódovanie**. Pri **vzorkovaní** sa odoberajú hodnoty (vzorky) v pravidelných intervaloch. Hustota odoberania vzoriek sa nazýva **vzorkovacia frekvencia**.

Pri **kvantovaní** ide o rozdelenie celého rozsahu hodnôt, ktoré môže nadobúdať analógová informácia na intervaly. Každý interval má potom priradenú svoju zástupnú hodnotu (kvantovanie podľa hodnoty). Kvantovaním zaradíme hodnoty analógovej informácie namerané vzorkovaním do príslušných intervalov.

Po zaradení do intervalov nastáva tretia fáza spracovania - **kódovanie**. Pri kódovaní sa uložia navzorkované a nakvantované hodnoty vo formáte vhodnom pre digitálnu techniku. V podobe slov zložených z núl a jednotiek - zakódujeme ich do binárnej podoby.