

Ludi môžeme rozdeliť do 10 skupín – na tých, ktorí rozumejú binárnej sústave a tých, ktorí jej nerozumejú.

Kódovanie

V živote sa stretávame s množstvom informácií. Pamätať si všetky nedokážeme. Preto ich určitým spôsobom kódujeme. Takým kódom je napr. aj naša adresa. Keď pozývame niekoho na návštevu, neopíšeme mu ako vyzerá náš dom, čo máme vysadené pred domom, aké auto nám stojí na dvore, ale zadáme jednoducho Nábrežná 42. Náš dom je zakódovaný ako ulica a číslo domu v ulici. Podobne sa v obchodoch používa čiarový kód na zakódovanie názvu a ceny tovaru. Kódovanie sa používa vtedy, keď vyjadrovanie bežnými slovami by bolo zdĺhavé alebo náročné na miesto a čas.

Kódovanie je proces, pri ktorom sa každému znaku alebo postupnosti znakov daného súboru znakov (vzorov) jednoznačne priradí znak alebo postupnosť znakov (obrazov) z iného súboru znakov.

Kódovanie je teda transformácia určitej informácie z jednej formy na druhú pomocou určitého postupu - algoritmu, ktorý je väčšinou verejne známy. Vo väčšine prípadov **účelom kódovania nie je utajenie informácie** (na rozdiel od šifrovania) **ale len jej iná forma zápisu** vybraná tak, **aby sa informácia dala čo najlepšie alebo najúspornejšie uchovať alebo preniesť** (Morseova abeceda, semafor, ASCII...).

Vďaka počítačom sa najčastejšie sa používa **kódovanie** údajov a informácií **do číselnej podoby**. Takémuto kódovaniu tiež hovoríme **digitalizácia**. Používa sa však aj nečíselné kódovanie. Typickým príkladom nečíselného kódovania z praxe je semafor červená = stoj, zelená = chod', Morseova abeceda.



A	.-	M	--	Y	-.--	6	-...-
B	-...-	N	-.	Z	--..	7	-...-
C	-.-.	O	---	Ä	.-.-	8	---..
D	--.	P	.-.	Ö	---.	9	-----
E	.	Q	--.-	Ů	..--	.	.-.-.-
F	..-.	R	.-.	Ch	----	,	---.-
G	--.	S	...	0	-----	?	..--..
H		T	-	1	.----	!	*--
I	..	U	..-	2	..----	:	---...
J	.-.-	V	...-	3	...--	"	.-.-.-
K	-.-	W	.-.	4	-	'	.-....
L	.-..	X	-..-	5		=	-...-

CVČ: Pomocou Morseovej abecedy zakódujte medzinárodné volanie o pomoc SOS a vaše meno.

Vieme, že počítač informácie kóduje pomocou rôznej kombinácie hodnôt bitov - najmenej jednotky informácie. Vieme tiež, že ak ho chceme používať na zbieranie, uchovávanie,

spracúvanie a šírenie informácií, musíme nájsť spôsob, ako informácie rôzneho typu preložiť do binárneho kódu.

Bits, pomocou ktorých kódujeme, sú do pamäťových buniek počítača ukladané po ôsmich, preto je výhodné na zakódovanie údajov použiť vždy taký počet bitov, ktorý je deliteľný ôsmimi (8,16,24,32).

Kódovanie čísel v počítači

Čísla sa vyjadrujú pomocou dvojkovej pozičnej sústavy. Na kódovanie čísel v počítačoch je najvýhodnejšie použiť jedno „slovo“ (word), t.j. taký počet bitov, ktoré počítač dokáže spracovať počas jedného taktu procesora (v súčasnosti 64 bit).

Kódovanie prirodzených čísel a nuly

Obrovskou výhodou je fakt, že každé číslo môžeme previesť do dvojkovej sústavy, ktorá používa iba cifry 0 a 1, čím pre každé číslo dostaneme jednoznačný zápis. Prirodzené čísla sú teda v počítači uložené v tzv. **priamom kóde**, čo je vlastne číslo prevedené do dvojkovej sústavy.

- Pri použití 1 Bajtu (8 bitov) môžeme zakódovať 256 možných hodnôt, t.j. čísla 0 – 255.
- Pri použití 2 Bajtov (16 bitov) môžeme zakódovať 2^{16} možných hodnôt, t.j. čísla 0 – 65 535.
- Pri použití 4 Bajtov (32 bitov) môžeme zakódovať 2^{32} možných hodnôt, t.j. čísla 0 – 4 294 967 295.
- Pri použití 1 slova (word – 64 bitov) môžeme zakódovať 2^{64} možných hodnôt, t.j. čísla 0 – 18 446 744 073 709 551 615.

Kódovanie celých čísel

Pri celých číslach je potrebné zohľadniť i znamienko. Našťastie sú znamienka len dve (+, -), preto ich môžeme zakódovať 1 bitom: 0 t.j. +, 1 t.j. -. Pri kódovaní celých čísel sa znamienko zakóduje vždy prvým bitom zľava. Napr. pri použití 1 Bajtu bude $(10011101)_2$ kód pre -29. Číslo $(1000010100111110)_2$ je -1342 (číslo 1342 v binárnej sústave je 10100111110, museli sme ho doplniť nulami na 16 cifier a na začiatok vyhradiť bite pre znamienko).

- Pri použití 1 Bajtu (8 bitov - 1 bit znamienko a 7 bitov hodnota), môžeme zakódovať čísla - 128 až + 127
- Pri použití 2 Bajtov (16 bitov -1 bit znamienko a 15 bitov hodnota), môžeme zakódovať čísla - 32 768 až + 32 767
- Pri použití 4 Bajtov (32 bitov - 1 bit znamienko a 31 bitov hodnota), môžeme zakódovať čísla - 2 147 483 648 až + 2 147 483 647
- Pri použití 8 Bajtov (64 bitov - 1 bit znamienko a 63 bitov hodnota), môžeme zakódovať čísla - 9 223 372 036 854 775 808 až + 9 223 372 036 854 775 807

Kódovanie reálnych čísel

Reálne čísla môžeme do počítača kódovať dvoma spôsobmi:

1. ako čísla s pevnou rádovou čiarkou

Pri tomto spôsobe kódovania je niekoľko bitov vyhradených pre celú časť čísla a niekoľko pre desatinnú časť čísla.

Ak pri nejakej operácii dostaneme väčší počet desatinných miest ako môžeme zakódovať pomocou vyhradeného počtu bitov, vtedy sa zvyšné miesta jednoducho odrežú a nebudú do pamäte počítača uložené.

$$(1001.101)_2 = 2^3 + 2^0 + 2^{-1} + 2^{-3} = 8 + 1 + 0.5 + 0.125 = (9.625)_{10}$$

Takéto čísla sa sčítajú a odčítajú rovnako ako celé čísla, ale násobenie a delenie je iné. Majú malý rozsah zobraziteľných hodnôt, preto sú nevhodné na vedecké výpočty. Majú menej problémov so zaokrúhľovaním, preto sa niekedy používajú na výpočty s peniazmi, napr.: 15,362 €, kde malý rozsah až tak nevadí

2. ako čísla s pohyblivou rádovou čiarkou

Niekoľko bitov je vyhradených pre hodnotu čísla (**mantisu**) a zvyšok je vyhradený pre exponent.

Napr.: kalkulačky, ktoré počítajú v desiatkovej sústave číslo 126,567 zobrazia ako $126567 \cdot 10^{-3}$. V našom prípade je mantisa 126567 a exponent -3. Obe tieto hodnoty sú uložené samozrejme v priamom kóde a majú 1 bit vyhradený pre znamienko.

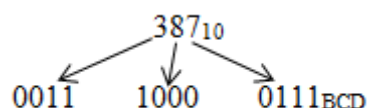
Tento spôsob zápisu čísla sa používa na zápis vedeckých výpočtov, nazývaný tiež semilogaritmický tvar alebo vedecká notácia. Napríklad hmotnosť slnka je asi $2 \cdot 10^{33}$ g, hmotnosť elektrónu je asi $9 \cdot 10^{-28}$ g. Rozdiel medzi nimi je 61 desiatkových rádov, čo je asi 202 dvojkových rádov. Na fyzikálny výpočet takú veľkú presnosť vôbec nevyužijeme lebo obe čísla poznáme s presnosťou ledva 10 cifier.

Napr.: predpokladajme, že počítač pracuje v desiatkovej sústave a máme zobrazit číslo 3213 v pohyblivej rádovej čiarkke, t.j. $3213 = 0.3213 \cdot 10^4$

BCD (Binary Coded Decimal) KÓD

Kód BCD je jedným z najčastejšie používaných kódov na reprezentovanie desiatkových čísel. Používal sa v starších procesoroch. Pre pomalšie počítače bol dobrým kompromisom dosť rýchleho prevodu z/do textového tvaru a rýchlosťou jednoduchých aritmetických výpočtov. Pri tomto kóde je každá desiatková číslica zakódovaná pomocou štyroch bitov.

Prevod dekadického čísla do kódu BCD:



Kódovanie textu v počítači

Uvedomme si, že ak nájdeme pre akýkoľvek typ informácie spôsob, ako ju zapisovať pomocou desiatkových čísel, získame tak zároveň aj spôsob jej zápisu v binárnom kóde.

Pred 5000 rokmi Féničania položili základy abecedy, v ktorej (na rozdiel od všetkých predchádzajúcich) je základom zápis hlásky, nie zápis celého slova. Aj pri prezentácii textov v počítači je takáto cesta správna – za základ zápisu textu zvolíme znak a nájdeme preň vhodné kódovanie.

Kódovanie znakov

Hľadáme spôsob, ako pomocou núl a jednotiek zapísať (zakódovať) písmená slovenskej abecedy (malé i veľké), a tiež číslce 0, 1, 2,..., 9 (pretože v texte sa niekedy používajú aj tie) a interpunkčné znamienka ako bodka, čiarka, úvodzovky... k týmto „viditeľným znakom“ musíme ešte pripojiť aj znaky „neviditeľné“ ako medzery, Enter, tab... Nesmieme však zabudnúť ani na české písmená Ř, ř, Ě, ě, ů, ů..., ktoré sa občas objavia napr. v menách a pod. znakov, ktoré sme spomenuli, je už viac než 128, čiže viac ako 2^7 , našťastie však menej ako 256, čiže 2^8 . Preto na kódovanie každého z nich musíme použiť 8 bitov (1 byte).

Priradenie binárnych kódov znakom sa nazýva **kódová tabuľka**: Windows 1250, ASCII, Unicode, UTF-8

Ako v nej čítame kódy? Predstavme si, že hľadáme napr. kód pre znak veľké Š. Tento znak nájdeme v stĺpci s nadpisom 128, v riadku s nadpisom 10, t.j. $(8A)_{16}$. Po prevode týchto čísel do dvojkovej sústavy dostávame $(10001010)_2$.

Opačne by sme postupovali takto: máme nájsť napr. znak, ktorého kód je $(01001101)_2$. Osem bitov rozdelíme na dve skupiny po 4 bity, čiže 0100 a 1101. Prvú štvoricu prevedieme do desiatkovej sústavy – výsledok určuje nadpis stĺpca. Druhá štvoricu prevedieme do desiatkovej sústavy, a tak určíme nadpis riadka $(1101)_2 = (13)_{10}$. Hľadaný znak je M. Znak $(C7)_{16} = C$.

	0	16	32	48	64	80	96	112	128	144	160	176	192	208	224	240
0				0	@	P	`	p	€			°	Ř	Đ	ř	đ
1			!	1	A	Q	a	q		'	~	±	Á	Ñ	á	ñ
2			"	2	B	R	b	r	,	'	~		Â	Ň	â	ň
3			#	3	C	S	c	s	f	"	Ł	ł	Ă	Ō	ă	ó
4			\$	4	D	T	d	t	„	”	¤		Ä	Ö	ä	ö
5			%	5	E	U	e	u	...	•	À	µ	Ĺ	Ő	í	ő
6			&	6	F	V	f	v	†	—	ı	¶	Č	Ö	č	ö
7			'	7	G	W	g	w	‡	—	§	·	Ç	×	ç	×
8			(	8	H	X	h	x	^	~	”	„	Č	Ř	č	ř
9			)	9	I	Y	i	y	‰	™	©	„	É	Ů	é	ů
10			*	:	J	Z	j	z	Š	š	Š	š	Ě	Ú	ě	ú
11			+	;	K	[	k	{	<	>	«	»	Ě	Ú	ě	ú
12			,	<	L	\	l		Š	š	¬	ˆ	Ě	Ú	ě	ú
13			-	=	M	]	m	}	Š	š	-	ˆ	Ě	Ú	ě	ú
14			.	>	N	^	n	~	Ž	ž	®	ˆ	Ě	Ú	ě	ú
15			/	?	O	_	o		Ž	ž	Ž	ž	Đ	ß	d'	'

Kódovanie slov

Pretože už vieme kódovať znaky, ľahko zakódujeme aj slová – jednoducho za seba zoradíme všetky kódy znakov. Nemusíme sa dokonca starať ani o oddelovanie slov, pretože aj medzera je znak a má svoj kód.

Tabuľka, ktorú sme uviedli, ilustruje iba jediné z mnohých možných kódovaní. Každý výrobca počítačov tento problém riešil iným spôsobom, preto existuje **viacero znakových kódov**.

Donedávna sme (podobne ako iné malé národy) považovali za samozrejmé, že počítače a programy sa s nami rozprávali po anglicky. Mohli by sme sa teda domnievať, že aspoň v anglicky hovoriacich krajinách bude kódovanie znakov jednotné a ustálené. Ale ani to nie je celkom pravda... Vznikali rôzne pokusy o jednotné kódovanie, najrozšírenejší z nich je kódovanie pomocou ASCII tabuľky a Unicode.

ASCII – American Standard Code for Information Interchange (Americký štandardný kód pre výmenu informácií)

Jeho nevýhodou je, že každý znak kóduje iba siedmimi bitmi, čo nám nevyhovuje, lebo možno zakódovať iba $2^7 = 128$ znakov (čo pre nás (slovenskú abecedu) nestačí).

Pre potreby ďalších jazykov sa používajú osembitové rozšírenia, ktoré obsahujú ďalších 128 kódov. Tak vznikli pre potreby jednotlivých jazykov rôzne kódové tabuľky, kde význam kódov nad 127 nie je jednoznačný.

PR: Veta "Ema má mamu." je v počítači pomocou ASCII uložená nasledovne:

69 109 97 32 109 113 32 109 97 109 117

Tabuľka základných znakov ASCII (128 znakov)

Code	Char	Code	Char	Code	Char	Code	Char	Code	Char	Code	Char
32	[space]	48	0	64	@	80	P	96	`	112	p
33	!	49	1	65	A	81	Q	97	a	113	q
34	"	50	2	66	B	82	R	98	b	114	r
35	#	51	3	67	C	83	S	99	c	115	s
36	\$	52	4	68	D	84	T	100	d	116	t
37	%	53	5	69	E	85	U	101	e	117	u
38	&	54	6	70	F	86	V	102	f	118	v
39	'	55	7	71	G	87	W	103	g	119	w
40	(	56	8	72	H	88	X	104	h	120	x
41	)	57	9	73	I	89	Y	105	i	121	y
42	*	58	:	74	J	90	Z	106	j	122	z
43	+	59	;	75	K	91	[	107	k	123	{
44	,	60	<	76	L	92	\	108	l	124	
45	-	61	=	77	M	93	]	109	m	125	}
46	.	62	>	78	N	94	^	110	n	126	~
47	/	63	?	79	O	95	_	111	o	127	[backspace]

Tabuľka znakov ASCII (256 znakov)

0		44	,	88	X	132	ä	176	⌘	220	░
1	☉	45	-	89	Y	133	å	177	⌘	221	▒
2	☼	46	.	90	Z	134	é	178	⌘	222	▓
3	♥	47	/	91	[	135	ç	179	⌘	223	█
4	♣	48	0	92	\	136	l	180	⌘	224	◼
5	♠	49	1	93	]	137	ë	181	⌘	225	▣
6	♣	50	2	94	^	138	ö	182	⌘	226	◼
7	•	51	3	95	_	139	õ	183	⌘	227	▤
8	▣	52	4	96	`	140	î	184	⌘	228	▥
9	◊	53	5	97	a	141	ž	185	⌘	229	▦
10	▣	54	6	98	b	142	Ä	186	⌘	230	▧
11	♂	55	7	99	c	143	Č	187	⌘	231	▨
12	♀	56	8	100	d	144	É	188	⌘	232	▩
13	♫	57	9	101	e	145	Ł	189	⌘	233	▪
14	♫	58	:	102	f	146	Í	190	⌘	234	▫
15	☼	59	;	103	g	147	ô	191	⌘	235	▬
16	▶	60	<	104	h	148	ö	192	⌘	236	▭
17	◀	61	=	105	i	149	Ë	193	⌘	237	▮
18	↑	62	>	106	j	150	ƒ	194	⌘	238	▯
19	!!	63	?	107	k	151	Š	195	⌘	239	▰
20	¶	64	@	108	l	152	ś	196	⌘	240	▱
21	§	65	A	109	m	153	Ö	197	⌘	241	▲
22	—	66	B	110	n	154	Ů	198	⌘	242	△
23	↑	67	C	111	o	155	Ť	199	⌘	243	▴
24	↑	68	D	112	p	156	ť	200	⌘	244	▵
25	↓	69	E	113	q	157	Ľ	201	⌘	245	▶
26	→	70	F	114	r	158	×	202	⌘	246	▷
27	←	71	G	115	s	159	ě	203	⌘	247	▸
28	└	72	H	116	t	160	á	204	⌘	248	▹
29	↔	73	I	117	u	161	í	205	⌘	249	►
30	▲	74	J	118	v	162	ó	206	⌘	250	▻
31	▼	75	K	119	w	163	ú	207	⌘	251	▼
32 (medzera)		76	L	120	x	164	À	208	⌘	252	▽
33	!	77	M	121	y	165	á	209	⌘	253	▾
34	„	78	N	122	z	166	Ž	210	⌘	254	▿
35	#	79	O	123	{	167	ž	211	⌘	255	▹
36	\$	80	P	124		168	Ě	212	⌘		
37	%	81	Q	125	}	169	č	213	⌘		
38	&	82	R	126	~	170	~	214	⌘		
39	•	83	S	127	△	171	ž	215	⌘		
40	(	84	T	128	Ç	172	Č	216	⌘		
41	)	85	U	129	ü	173	š	217	⌘		
42	*	86	V	130	é	174	«	218	⌘		
43	+	87	W	131	â	175	»	219	⌘		

Vzt'ah tabuľky Windows CP1250 a ASCII tabuľky (128 znakov): ak spočítame súčet čísel v záhlaví tabuľky pre každé písmeno v tabuľke Windows1250 dostávame kód pre dané písmeno v tabuľke ASCII. Napr. písmeno X je v tabuľke Windows 1250 v stĺpci označenom 80 a riadku označenom 8: $80 + 8 = 88$, čo je kód pre X v tabuľke ASCII.

Unicode

Druhý zdroj problémov v kódovaní textu (a teda napr. v jeho prenositeľnosť kódovania z krajiny do krajiny) sú národné abecedy. Dnes by sme sa už sotva uspokojili s digitálnymi dokumentmi bez mäčkových a dĺžňov. Veď počítač a nové technológie majú slúžiť nám, nie my im. Vedia to aj najväčší výrobcovia softvéru a svoje nové programy robia tak, aby komunikovali so zákazníkom v jeho jazyku a aby správne pracovali s textami vyrobenými v jeho jazyku.

Novým pokusom o univerzálne riešenie kódovacích problémov je jednotné kódovanie Unicode. Každý znak sa v ňom zapisuje pomocou 16 bitov (2 B), takže dovoľuje používať až $2^{16} = 65\,536$ rôznych znakov zo všetkých bežných jazykov.

Tento počet znakov umožňuje zakódovať znaky všetkých bežne používaných abecied pomocou jednej medzinárodnej tabuľky. Toto kódovanie zabezpečuje, že ten istý znak má rovnaký kód v každej krajine aj na každom type počítača. Nevýhodou tohto kódovania je, že každý znak je kódovaný 16 bitmi, a teda zaberá viac pamäte ako kód ASCII.

UTF-8

Istým vylepšením kódovania Unicode je **kódovanie UTF-8**. V tomto kódovaní je prvých 128 znakov tabuľky ASCII (tieto sú pre všetky krajiny rovnaké), zakódovaných pomocou 8 bitov a zvyšné znaky sú zakódované 16, 24, 32, 40 až 48 bitmi. Toto kódovanie je výhodné pre americky hovoriace krajiny a krajiny, v ktorých väčšinu znakov textu tvorí prvých 128 znakov tabuľky ASCII.