

Reprezentácia údajov v počítači, Čísla, znaky, Komprimácia, Šifry.

□ Počítače uchovávajú všetky údaje pomocou čísel, spôsob resp. proces akým sú nečíselné údaje reprezentované pomocou čísel. sa nazýva **digitalizácia dát** a spôsob, akým sa údaje digitalizujú, voláme **kódovanie údajov**.

□ **Kódovanie** je transformácia určitej informácie z jednej formy na druhú pomocou určitého algoritmu s použitím parametra. Ak je parameter verejný - ide o kódovanie, ak je utajený - **šifrovanie**.

Kódovať však môžeme údaje i do nečíselnej podoby.

Typickým príkladom nečíselného kódovania je Morseova abeceda alebo Braillovo písmo pre nevidiacich.

□ Kódovanie čísel v počítači.

Všetky údaje v počítači sú kódované pomocou rôznej kombinácie hodnôt bitov - najmenej jednotky informácie. Každý z bitov môže nadobúdať iba dve rôzne hodnoty 0 a 1.

Tieto bity sú však do pamäťových buniek počítača ukladané po osmiciach, preto je výhodné na zakódovanie údajov použiť vždy taký počet bitov, ktorý je deliteľný ôsmimi (8,16,24,32). Čím väčší počet bitov použijeme, tým väčší rozsah čísel môžeme použiť. Napríklad pomocou 8 bitov dostaneme $2^8 = 256$ rôznych kombinácií núl a jednotiek.

Pomocou 8 bitov teda môžeme zakódovať napríklad čísla od 0 do 255 alebo čísla od -128 do 127 v prípade, že potrebujeme i záporné čísla.

Na kódovanie čísel v počítačoch je najvýhodnejšie použiť jedno „slovo“ (**Word**), t.j. taký počet bitov, ktoré počítač dokáže spracovať počas jednej operácie (jedného taktu procesora). Najmodernejšie počítače dnes používajú 64 bitové slovo, teda dokážu spracovať 64 bitov pri jednej operácii.

Kódovanie prirodzených čísel a nuly.

Výhodou je, že každé číslo môžeme previesť do dvojkovej sústavy, 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 jedného Bajtu (8 bitov) môžeme zakódovať 256 možných hodnôt, t.j. hodnoty 0 až 255

□ pri použití 2 Bajtov (16 bitov) hodnoty 0 až 65535

□ pri použití 4 Bajtov (32 bitov) hodnoty 0 až 4 294 967 295

□ pri použití 1 slova uvažovaného súčasného počítača (64 bitov) hodnoty 0 až 18 446 744 073 709 551 615, čo sú už astronomické čísla

Kódovanie celých čísel.

Pri celých číslach je potrebné zohľadniť i znamienko. ktoré sa zakóduje 1 bitom (0 = +, 1 = -). Pri kódovaní celých čísel sa znamienko zakóduje vždy prvým bitom zľava.

Napr. pri použití 1 Bajtu bude 10011101 kód pre -29.

□ pri použití 1 Bajtu (8 bitov - 1 bit znamienko a 7 bitov hodnota), môžeme zakódovať hodnoty -128 až +127

□ pri použití 2 Bajtov (16 bitov - 1 bit znamienko a 15 bitov hodnota), môžeme zakódovať hodnoty -32 768 až +32 767

□ pri použití 4 Bajtov (32 bitov - 1 bit znamienko a 31 bitov hodnota), môžeme zakódovať hodnoty -2 147 483 648 až +2 147 483 647

□ pri použití 4 Bajtov (64 bitov - 1 bit znamienko a 63 bitov hodnota), môžeme zakódovať hodnoty -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 (tento spôsob sa väčšinou používa na uloženie meny napr.: 1,33€). Pri tomto spôsobe 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.

2. Ako čísla s pohyblivou rádovou čiarkou – tu je vyhradených niekoľko bitov pre hodnotu čísla (**mantisu**) a zvyšok je vyhradený pre **exponent**.

Napr.: číslo 126,567 je uložené ako 126567.10⁻³.

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.

Kódovanie znakov.

Na rozdiel od čísel, znaky textu nevieme previesť do dvojkovej sústavy, preto bolo potrebné vymyslieť iný spôsob ako jednoznačne priradiť určitému znaku práve jednu kombináciu núl a jednotiek, ktorá tento znak v počítači bude reprezentovať. Keďže neexistuje žiadny univerzálny spôsob ako to urobiť, každý výrobca počítačov tento problém riešil iným spôsobom, preto existuje viacero znakových kódov. Poriadok do tohto chaosu sa snažil zaviesť americký úrad pre normalizáciu, ktorý vyhlásil jeden spôsob, ktorý by mali všetci používať. Tento spôsob kódovania sa volá ASCII - American Standard Code for Information Interchange .

Kód ASCII.

Tento štandard hovorí, že na zakódovanie každého znaku sa použije 8 bitov. Čo umožňuje definovať kód pre 256 znakov. Pričom prvá polovica znakov bude pre všetky krajiny rovnaká a zvyšných 128 znakov sa pre každú krajinu stanovil podľa ich potrieb. Prvej polovici znakov preto prislúchajú čísla v dvojkovej sústave, ktoré po prevode do desiatkovej sústavy majú hodnoty 0 -127, a týmto hodnotám prislúchajú tieto znaky:

ASCII Table

Character	Decimal Number	Binary Number	Character	Decimal Number	Binary Number
blank space	32	0010 0000	^	94	0101 1110
!	33	0010 0001	-	95	0101 1111
"	34	0010 0010	`	96	0110 0000
#	35	0010 0011	a	97	0110 0001
\$	36	0010 0100	b	98	0110 0010
A	65	0100 0001	c	99	0110 0011
B	66	0100 0010	d	100	0110 0100
C	67	0100 0011	e	101	0110 0101
D	68	0100 0100	f	102	0110 0110
E	69	0100 0101	g	103	0110 0111
F	70	0100 0110	h	104	0110 1000
G	71	0100 0111	i	105	0110 1001
H	72	0100 1000	j	106	0110 1010
I	73	0100 1001	k	107	0110 1011
J	74	0100 1010	l	108	0110 1100
K	75	0100 1011	m	109	0110 1101
L	76	0100 1100	n	110	0110 1110
M	77	0100 1101	o	111	0110 1111
N	78	0100 1110	p	112	0111 0000
O	79	0100 1111	q	113	0111 0001
P	80	0101 0000	r	114	0111 0010
Q	81	0101 0001	s	115	0111 0011
R	82	0101 0010	t	116	0111 0100
S	83	0101 0011	u	117	0111 0101
T	84	0101 0100	v	118	0111 0110
U	85	0101 0101	w	119	0111 0111
V	86	0101 0110	x	120	0111 1000
W	87	0101 0111	y	121	0111 1001
X	88	0101 1000	z	122	0111 1010

Pre Slovensko druhú polovicu tabuľky stanovuje medzinárodná norma **ISO 8859-2**, ktorá sa tiež nazýva **Latin 2**.
Kódovanie Unicode

I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143
I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159
nbsp	À	Á	Â	Ã	Ä	Å	Š	Ś	Š	Ť	Ž	shy	Ž	Ž	
160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175
°	à	á	â	ã	ä	å	š	ś	š	ť	ž	”	ž	ž	
176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191
Ŕ	Á	Â	Ã	Ä	Å	Ł	Ć	Ç	Č	É	Ê	Ë	Ě	Í	Î
192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207
Đ	Ñ	Ň	Ó	Ô	Õ	Ö	×	Ř	Ů	Ú	Ů	Ü	Ý	Ť	ß
208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223
í	á	â	ã	ä	å	ł	ć	ç	č	é	ê	ë	ě	í	î
224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239
đ	ñ	ň	ó	ô	õ	ö	÷	ř	ů	ú	ů	ü	ý	ť	·
240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255

Toto kódovanie používa 16 bitov na zakódovanie jedného znaku, čo umožňuje zakódovať 65536 možných znakov. Tento počet znakov umožňuje zakódovať znaky všetkých abecied pomocou jednej medzinárodnej tabuľky. Toto kódovanie zabezpečuje, že ten istý znak má rovnaký kód v každej krajine i na 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. Istým vylepšením tohto kódovania 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.

Kompresia – komprimácia údajov

Kompresia (stlačenie, zhustenie) je snaha pomocou rôznych postupov zapísať dáta tak, aby zaberali menej miesta na záznamovom médiu. Zapísané dáta sa však musia dať previesť späť do pôvodného stavu alebo musia aspoň zachovať obsahový význam uložených dát (stratová kompresia).

Kompresia šetrí miesto

To v praxi znamená, že na rovnaké záznamové médium sa nám zmestí viac údajov. Napríklad ak pesničky skomprimujeme do formátu MP3, na pamäťové médium sa nám ich zmestí oveľa viac než pesničiek, ktoré nie sú komprimované. Ak obrázky skomprimujeme do formátu JPEG, oveľa viac obrázkov sa zmestí ako keby boli uložené vo formáte BMP.

Kompresia šetrí čas

Ak posielame komprimovaný súbor cez sieť, **prenos bude rýchlejší**, ak údaje pred prenosom najprv skomprimujeme. Napríklad ak posielame pesničky vo formáte MP3, jej prenos bude 12-krát rýchlejší ako pri nekomprimovanej pesničke. Rovnako, ak umiestnime obrázok vo formáte JPEG na web stránku, tak jeho načítanie v prehliadači bude rýchlejšie ako keby sme ho tam umiestnili vo formáte BMP.

Kompresné formáty, kompresný pomer

Okrem obrázkov a hudby môžeme komprimovať akékoľvek dáta.

Na tento účel slúžia kompresné formáty ako napríklad **ZIP**, **GZIP**, **RAR**, **7-ZIP**....

O úspešnosti použitia kompresie hovorí **kompresný pomer**, čo je pomer skomprimovaných dát ku neskomprimovaným ak sa podarí súbor zmenšiť, je kompresný pomer vždy menší ako 1.

Typy kompresie

Stratová kompresia(lossy)

Odstráni nepodstatné údaje z dát (nepočuteľné zvuky, neviditeľné body, splývanie farieb, ...), zachová sa iba významový obsah údajov, údaje nie sú 100% rekonštruovateľné. (MP3.JPEG,...)

Bezstratová kompresia(lossless)

Zmenší údaje tak, aby sa potom dali dať späť do pôvodného stavu, zachová sa pôvodný údaj bez akejkoľvek straty, údaje sú 100% rekonštruovateľné.

(GIF, PNG, ...)

Kompresné algoritmy

Medzi najznámejšie kompresné algoritmy patria Huffmanov kód, Run-Length Encoding (RLE) a LZW.

Huffmanov kód

Veľmi účinná kompresia používaná aj formátmi JPEG a MPEG.

Najčastejšie používané znaky sa kódujú menším počtom bitov ako normálne. Najskôr sa vytvorí tabuľka počtu výskytov všetkých znakov. Potom sú symboly usporiadané od najčastejšie použitých po najmenej často použité. Častejším symbolom sa potom priradí menší počet bitov ako menej častým.

Run-Length Encoding (RLE)

Kompresia vhodná na použitie v reálnom čase, striedavo sa zapisuje počet jednotiek a počet núl vstupu, ktoré sa zapíšu v dvojkovej sústave.

LZW

Používa sa napríklad vo formáte GIF.

Vytvára dynamický slovník ľubovoľne spájaných znakov, tzv. „slov“, ktorým sa priradujú čísla podľa ich prvého výskytu. Na začiatku sú v slovníku iba slová s dĺžkou jedna. Potom pribúdajú slová s dĺžkou dva a potom dlhšie a dlhšie. To znamená, že zo začiatku sú nahrádzané iba malé postupnosti, ale neskôr sa jedným číslom môže nahradiť i dlhé slovo.

Komprimačné formáty:

- a) Audio: mp3, wma, wav, flac
- b) Video: avi, wmv, mpeg – 1,2,4, theora
- c) Obrázky: jpeg, gif, png,

Kodek (kóder-dekóder) sa štandardne používa na označenie programu, ktorý dokáže kódovať a dekódovať určitý formát dát alebo signálov. Kodeky dodržiujúce jeden konkrétny štandard sa obvykle líšia kvalitou výsledného audia alebo videa a tiež podporou pre niektoré špecifické vlastnosti štandardu. Nie vždy je však kóder a dekóder jeden program a nie vždy je nutné používať oba. Na prehrávanie postačuje dekóder, pričom na vytváranie súborov je potrebný kóder.

Najznámejšie kodeky:

video: Divx Pro Codec, Xvid, Windows Media Video 9 series, libtheora, libavacodec, x264 audio: FLAC, Apple Lossless, Windows Media Audio 9 series, FAAC(AAC), LAME(MP3)