

Teorie informace

Jednotky informace

Bit (b)

Bit je základní a nejmenší jednotkou informace. Bit může nabývat pouze jednu ze dvou hodnot (jeden ze dvou stavů), které se dají interpretovat například jako 1 / 0, true / false, atd. Skupina 8 bitů tvoří jeden byte.

Byte (B, bajt)

Byte je jednotka množství dat v informatice. Skládá se z 8 bitů, tudíž může reprezentovat například číslo od 0 do 255 nebo jeden znak. Jeden byte je v obvykle nejmenší objem dat, se kterým dokáže procesor přímo pracovat.

Dříve byte označoval skupinu 1-6 bitů, dnes je standard 8 bitů.

Ve frankofonních zemích se někdy pro bajt používá značka o jako oktet (octet).

Násobnosti a převody binárních jednotek

V prosinci 1998 IEC (International Electrotechnical Commission) vytvořila dodatek k normě IEC 60027-2, ve kterém zavedla pro počítačové jednotky nový systém označování násobků. V tomto systému bylo pro původní „velké kilo“ = 1024 B navrženo označení kibibajt a značka KiB, zatímco jednotka kilobajt (se značkou kB) označuje 1000 B, tak jak je obvyklé v soustavě SI.

Binární předpona

Binární předpona je předpona jednotky vyjadřující násobek mocniny 2, podobně jako klasické předpony z SI vyjadřují násobek 10. Binární předpona vyjadřuje vždy o něco vyšší hodnotu než předpona klasická.

Dvojkový řád $n: 2^n$	Desítkový řád $k: 10^k$	Značka	Celý název	Hodnota
2^{10}		KiB	kibibajt	1 024
2^{20}		MiB	mebibajt	1 048 576
2^{30}		GiB	gibibajt	1 073 741 824
2^{40}		TiB	tebibajt	1 099 511 627 776
2^{50}		PiB	pebibajt	1 125 899 906 842 624
2^{60}		EiB	exbibajt	1 152 921 504 606 846 976
2^{70}		ZiB	zibibajt	1 180 591 620 717 411 303 424

Dvojkový řád $n: 2^n$	Desítkový řád $k: 10^k$	Značka	Celý název	Hodnota
2^{80}		YiB	yobibajt	1 208 925 819 614 629 174 706 176
	10^3	KB	kilobajt	1 000
	10^6	MB	megabajt	1 000 000
	10^9	GB	gigabajt	1 000 000 000
	10^{12}	TB	terabajt	1 000 000 000 000
	10^{15}	PB	petabajt	1 000 000 000 000 000
	10^{18}	EB	exabajt	1 000 000 000 000 000 000
	10^{21}	ZB	zettabajt	1 000 000 000 000 000 000 000
	10^{24}	YB	yottabajt	1 000 000 000 000 000 000 000 000

Přehled násobných jednotek

Jednotka	Značka	B	kB	KiB	MB	MiB	GB	GiB	TB	TiB
Kilobajt	kB	1000	1	~0,9766						
Kibibajt	KiB	1024	1,024	1						
Megabajt	MB	1 000 000	1000	~976,6	1	~0,9537				
Mebibajt	MiB	1 048 576	~1048,6	1024	1,049	1				
Gigabajt	GB	10^9	1 000 000 976 562,5	1000	953,7	1	~0,9313			
Gibibajt	GiB	$\sim 1,074 \cdot 10^9$	~1 073 742	1 048 576	~1073,7	1024	1,074	1		
Terabajt	TB	10^{12}	10^9	$\sim 0,9766 \cdot 10^9$	1 000 000	~953 674,3	1000	931,3	1	~0,9095
Tebibajt	TiB	$\sim 1,1 \cdot 10^{12}$	$\sim 1,1 \cdot 10^9$	$\sim 1,074 \cdot 10^9$	~1 099 512	1 048 576	~1099,5	1024	~1,1	1

Binární a hexadecimální soustava, převody čísel BIN-HEX-DEC

Binární (dvojková) soustava

Binární soustava je číselná soustava, která používá pouze dva symboly: 0 a 1. Používá se ve všech moderních digitálních počítačích, neboť její dva symboly (0 a 1) odpovídají dvěma jednoduše rozdělitelným stavům elektrického obvodu (vypnuto a zapnuto), popřípadě nepravdivosti či pravdivosti výroku. Číslo zapsané v dvojkové soustavě se nazývá binární číslo.

Příklad čísla, zapsaného v binární soustavě				
Rozepsané číslo	1	0	1	0
Násobeno	2^3	2^2	2^1	2^0
Rozepsaný násobek	8	4	2	1
V desítkové soustavě	10			

Hexadecimální (šestnáctková) soustava

Hexadecimální čísla se zapisují pomocí číslic '0', '1', '2', '3', '4', '5', '6', '7', '8' a '9' a písmen 'A', 'B', 'C', 'D', 'E' a 'F', přičemž písmena 'A'-'F' reprezentují cifry s hodnotou 10-15). Čísla v tomto zápisu se obvykle označují písmenem H připojeným k číslu v dolním indexu. Např. $3F7_H$ reprezentuje hodnotu, které v desítkové soustavě odpovídá číslu $3 \times 16^2 + 15 \times 16^1 + 7 \times 16^0 = 1015$.

Díky jednoduchému vzájemnému převodu mezi šestnáctkovou a dvojkovou soustavou, se hexadecimální zápis čísel často používá v oblasti informatiky, například pro adresy v operační paměti počítače.

Převody čísel

Převod šestnáctkových čísel na dvojkové

Převod čísla z hexadecimální soustavy do soustavy dvojkové (binární) je usnadněn díky tomu, že číslo 16 je mocninou čísla 2 ($2^4 = 16$). Postup převodu je následovný. Rozdělíme byte reprezentovaný dvěma šestnáctkovými čísly na nibbly (1/2 bytu – 1 písmeno) a každý nibbl převedeme pomocí následující tabulky do jeho dvojkové (binární) reprezentace.

Šestnáctkové číslo	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
Dekadické číslo	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Binární číslo	0000	0001	0010	0011	0100	0101	0110	0111	1000	1001	1010	1011	1100	1101	1110	1111

Po převodu opět spojíme nibbly (teď již ve dvojkové (binární) reprezentaci) do jednoho bytu (teď již ve dvojkové (binární) reprezentaci).

Převod z dvojkové soustavy do šestnáctkové

Algoritmus převodu je přesně opačný, než u převádění HEX do binární soustavy. Nejprve rozdělíme byty na nibbly, které pomocí výše uvedené tabulky převedeme na jednotlivé číslice v hexadecimální soustavě, které spojíme dohromady.

Převod celých desítkových čísel na šestnáctkové

Celá desítková čísla můžeme převádět na šestnáctková například pomocí postupného dělení šestnáctí a sepisování zbytku po dělení.

Mějme například číslo $x = (15119)_{10}$ v dekadické soustavě. Převod provádíme tak, že číslo x dělíme šestnáctí a výsledek (podíl) píšeme v celých číslech. Při dělení vzniká zbytek, který si napíšeme. Vzniklý podíl opětovně dělíme šestnáctí a zbytek zapisujeme, dokud nedostaneme nulu. Když přečteme zbytky v obráceném pořadí jako šestnáctkové číslice, dostáváme šestnáctkové číslo:

15119 / 16	= 944	zbytek 15	(F)16
------------	-------	-----------	-------

944 / 16	= 59	zbytek 0	(0)16
59 / 16	= 3	zbytek 11	(B)16
3 / 16	= 0	zbytek 3	(3)16

Když přepíšeme zbytky v opačném pořadí, dostaneme šestnáctkové číslo 3B0F₁₆.

Znakové sady (ASCII, Windows-1250, ISO-8859-2, UTF-8)

Znaková sada nebo také kódová stránka v informační technologii je kód, který každému znaku (abecedy) přiřazuje určité číslo (bajt, sekvenci elektrických pulzů ap.). Výraz sada odráží to, že obsahuje kód pro určitou množinu znaků (např. obsahuje latinku bez nebo s určitými diakritickými znaménky), kód všech existujících znaků zahrnuje standard Unicode využívaný i moderním kódováním UTF-8. Převod textu do posloupnosti (sekvence) čísel a zpět slouží pro ukládání textu v počítači, jeho přenos telekomunikačními sítěmi apod.

ASCII

ASCII je anglická zkratka pro American Standard Code for Information Interchange („americký standardní kód pro výměnu informací“). V podstatě jde o kódovou tabulku, která definuje znaky anglické abecedy a jiné znaky používané v informatice. Jde o historicky nejúspěšnější znakovou sadu, z které vychází většina současných standardů pro kódování textu přinejmenším v euro-americké zóně. Veškeré moderní znakové sady jsou s ASCII interoperabilní v rozsahu 0-127, tedy další znaky jsou v rozsahu 128-256.

Prvních 32 znaků ASCII se nazývá kontrolní, a význam mnoha z nich je spíše historický, neboť ASCII původně ovládalo dálnopisy. V současnosti se hlavně používají tyto:

- SPACE
- NEWLINE FEED
 - posun o sloupec dolů, s výjimkou textových souborů UNIX, kde značí nový řádek
- CARRIAGE RETURN
 - návrat na začátek řádku
- BACKSPACE
- ESCAPE

[ASCII tabulka](#)

Windows-1250

Windows-1250, někdy také CP-1250, je znaková sada používaná operačním systémem Microsoft Windows pro reprezentaci textů ve středoevropských jazycích používajících latinku – např. pro albánštinu, chorvatštinu, češtinu, polštinu, rumunštinu, slovenštinu a maďarštinu. Windows-1250 se podobá sadě ISO 8859-2 — obsahuje všechny její tisknutelné znaky (a ještě několik navíc), ale několik z nich je na jiných místech (na rozdíl od Windows-1252, kde jsou všechny tisknutelné znaky na stejném místě jako v ISO 8859-1).

ISO-8859-2

ISO-8859-2, nebo také Latin-2 (pro odlišení od znakové sady CP852, která se používá v MS-DOS, a která se označuje jako Latin 2 se dá sada ISO-8859-2 označit jako ISO Latin 2 a CP852 jako PC Latin 2 nebo IBM Latin 2) je znaková sada, která obsahuje 191 znaků latinky kódovaných pomocí 8 bitů a byla poprvé publikována v roce 1987.

Srovnání Windows-1250 a ISO-8859-2

ISO 8859-2 se velmi podobá znakové sadě Windows-1250, která vznikla později. Kódování Windows-1250 zavedené firmou Microsoft obsahuje všechny tisknutelné znaky ISO 8859-2, ale některé z nich na jiných místech. V češtině je tento rozdíl nejmarkantnější u písmen Š, Ť a Ž.

UTF-8

UTF-8 je nejčastějším zápisem znakové sady Unicode. Unicode je narozdíl od výše zmíněných znakových sad určeno pro všechny světové jazyky najednou, protože znakům přiřazuje čísla až do 16 miliónů (zapisuje se většinou dvěma bajty). Zápis UTF-8 byl navržen pro zpětnou kompatibilitu s ASCII, se kterým tak má totožný způsob kódování 1-bajtových znaků. UTF-8 používá proměnnou délku znaku, a to od 1 bajtu až do 6 bajtů.

V dnešní době je UTF-8 dominantní znakovou sadu používanou na World Wide Web. (Používá ji více než 50% všech webových stránek. World Wide Web Consortium (W3C) doporučuje používat primárně UTF-8 a Internet Mail Consortium (IMC) doporučuje, aby všichni e-mailoví klienti uměli zobrazit a vytvářet e-maily kódované v UTF-8.).

Problémem UTF-8 je, že nelze přímo najít například 5. znak. Je potřeba postupně projít celý řetězen a počítat znaky, což dělá UTF-8 nepraktické pro reprezentaci řetězců při programování. Mnoho jazyků proto používá jiné kódování, například UTF-16, což tento problém ovšem řeší jen částečně. (Také je zde problém s definicí znaku kvůli kombinujícím znakům Unicode).

From:

<https://wiki.gml.cz/> - GMLWiki

Permanent link:

<https://wiki.gml.cz/doku.php/informatika:maturita:2a?rev=1600365992>

Last update: **17. 09. 2020, 20.06**

