

# Prezentácia (kódovanie) údajov čísel a znakov v počítači, ASCII kód, UNICODE, UTF-8 :)

## Kódovanie čísel v počítači

Všetky údaje v počítači sú kódované pomocou rôznej kombinácie hodnôt bitov - najmenšej jednotky informácie. Každý z bitov môže nadobúdať iba dve rôzne hodnoty 0 a 1. Tieto bity sú však do pamäťových buniek počítača ukladané po osmiaciach, 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“, 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

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 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 moderného počítača (64 bitov) hodnoty 0 až 18 446 744 073 709 551 615, čo sú už astronomické čísla.

V niektorých prípadoch (napríklad pri prenose) je vhodnejšie použiť iný kód ako je zápis čísla v dvojkovej sústave. Jedným z takýchto kódov je kód BCD.

## Kódovanie celých čísel

Pri celých číslach je potrebné zohľadniť aj znamienko. Našťastie znamienka sú len dve (+, -), preto ich môžeme zakódovať 1 bitom (0 = +, 1 = -). Pri kódovaní celých čísel sa znamienko zakóduje vždy prvým bitom zľava. Napríklad 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.

Celé čísla sú v pamäti počítača ukladané v priamom kóde, ale pre počítanie s nimi je vhodnejší doplnkový kód, ktorý získame z inverzného kódu.

## Kódovanie reálnych čísel

Reálne čísla môžeme do počítača kódovať dvoma spôsobmi:

Ako **čísla s pevnou rádovou čiarkou** (tento spôsob sa väčšinou používa na uloženie meny napríklad.: 24,50 €). Pri tomto spôsobe je niekoľko bitov vyhradených pre celú časť čísla a niekoľko pre desatinnú časť čísla. Ak nám 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.

Ako **čísla s pohyblivou rádovou čiarkou**. Tu je vyhradených niekoľko bitov pre hodnotu čísla (mantisu) a zvyšok je vyhradený pre exponent. Napríklad číslo 126,567 je uložené 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.

## 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 Americký štandardný kód pre výmenu informácií (American standard code for information interchange) - **ASCII**.

Tento štandard hovorí, že na zakódovanie každého znaku sa použije 8 bitov (pôvodne sa používalo len 7 bitov, ôsmy bol kontrolný). 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. Tento spôsob vniesol do kódovania znakov neuveriteľný chaos.

Preto sa vymyslel nový spôsob kódovania **UNICODE**. Toto kódovanie používa 16 bitov na zakódovanie jedného znaku, čo umožňuje zakódovať 65 536 možných znakov. Tento počet znakov umožňuje zakódovať znaky všetkých bežne používaných abecied pomocou jednej medzinárodnej tabuľky. Tento spôsob kódovania používa i kancelársky balík MS Office. Toto kódovanie zabezpečuje, že ten istý znak má rovnaký kód v každej krajine aj 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 ako kód ASCII. 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. Toto kódovanie je výhodné pre americky hovoriace krajiny a krajiny, v ktorých väčšinu znakov textu tvorí prvých 128 znakov tabuľky ASCII.