

## Računalniška predstavitev informacij

**S tem, ko smo definirali enoto informacije, smo dobili tudi praktično metodo za zapis informacije.** Bit ne predstavlja samo enote za merjenje informacije, ampak je tudi znak za zapis (oz.kodiranje) informacije. Bit ima lahko vrednosti 0 ali 1 in to sta elementa dvojiškega številskega sestava. (ugotavljali smo igralno barvo neke karte; to lahko zapišemo kot zaporedje znakov odgovorov na vprašanja; ali je črna -> ali je srce; [01]; vse kombinacije so 00, 01, 10, 11., s čimer smo kodirali vse možne izide.)

Računalniški zapis je v zelo tesni povezavi s teorijo informacije. Vsakovrstne informacije lahko razbijemo v zaporedja dvojiških števil, pri čemer mora biti jasno, kaj predstavlja posamičen bit (odgovor na katero vprašanje)!

Na tehničnem nivoju računalnik obdeluje in shranjuje informacije v obliki električnih impulzov! Teorija informacij podaja teoretično podlago tega procesa. Vrednost posameznega bita (0,1) je mogoče predstaviti z električnimi signali. Enica je predstavljena s tokovnim ali napetostnim sunkom, medtem ko njuna odsotnost pomeni ničlo.

### **A. Dva splošna načina predstavitve informacij: analogni (zvezni) in digitalni (diskretni)**

Latinski izvor besed: digitalni (sestavljeno iz znakov) in analogni (soroden fizikalni količini);

Matematična opredelitev razlike med digitalnim in analognim: Zaloga vrednosti diskretne funkcije je števna (da se prešteti vse signale), zaloga vrednosti zveznih funkcij pa so realna števila (množica z neskončnim številom elementov).

Računalnik po definiciji obdeluje samo binarne podatke – torej diskretne. Vendar to ne pomeni, da ne moremo predstaviti in obdelati analognih podatkov (potrebno jih je transformirati). Informacija je v računalniku vedno predstavljena s pomočjo samo dveh znakov – 0 in 1.

### **B. Predstavitev numeričnih informacij v bitnem zapisu**

Ksako lahko predstavimo neko število v računalniški obliki? V bitih, kar pomeni v dvojiškem številskega sistemu.

Desetiški številski sistem ali sistem z osnovo 10 je tisti, na katerega smo najbolj navajeni in s katerim delamo v vsakdanjem življenju (sestoji se iz števil 0..9). Če je število večje od 10, pa uporabljamo idejo o položaju cifre glede na decimalno vejico (bolj kot je levo od vejice, večji je utežni faktor cifre).

**Obstajajo tudi drugi številski sistemi**, pri čemer se v zvezi z računalniško tehnologijo uporabljajo dvojiški (bin), osmiški (oct) in šestnajstiški (hex) številski sistem. Osnovna števila so ustrezno 0 in 1 (bin), 0..7 (oct) in 0-9, a, b, c, d, e in f (hex).

Vsako pozitivno število se da zapisati na natanko en način v obliki:

$$N = x_n B^n + x_{n-1} B^{n-1} + \dots + x_0 B^0 + x^{-1} B^{-1} + \dots + x_{-m} B^{-m}$$

B: številna osnova

Indeks teče od  $-m$  do  $n$ :

- $m$  do  $0$ : število cifer desno od decimalne vejice;
- $1$  do  $n$ : število cifer levo od decimalne vejice;
- konkreten indeks predstavlja utežni faktor (oz. potenco)
- $x$  je med  $0$  in  $B-1$

$$4362,24 = 4 \cdot 10^3 + 3 \cdot 10^2 + 6 \cdot 10^1 + 2 \cdot 10^0 + 2 \cdot 10^{-1} + 4 \cdot 10^{-4} = 4000 + 300 + 60 + 2 + 0,2 + 0,04$$

### C. Pretvarjanje številskih sistemov

1.) iz desetiškega sistema v druge sisteme

Pomagamo si s Hornerjevo formulo pri pretvarjanju iz desetiškega sistema v druge sisteme:

Postopek:

- Celi del desetiškega števila pretvorimo v poljubni sistem z osnovo  $B$  tako, da ga delimo z osnovo  $B$ . Ostanke pri deljenju označujemo z  $a_i$  ( $i=0..n$ ). Postopek deljenja ponavljamo (cele dele delimo z  $B$ ) dokler ne pridemo do zadnjega ostanka  $a_n$ , ko postane količnik enak  $0$ . Iz dobljenih ostankov zapišemo celi del števila v sistemu  $B$ :  $a_n a_{n-1} \dots a_0 [b]$
- Neceli del števila pretvorimo v sistem z osnovo  $B$  tako, da neceli del množimo z osnovo  $B$ . Pri vsakem množenju odvezamo dobljeni cel del števila, ga zapišemo kot koeficient  $a_i$  in ostanek ponovno pomnožimo. Postopek ponavljamo, dokler a) z množenjem decimalna mesta niso enaka nič, ali b) dokler ne dobimo zadovoljivega števila decimalnih mest.

npr.  $131 [10] = 10000011 [2] = 203 [8] = 83 [16]$

- zadnji ostanek je  $a_n$  (tam je količnik  $0$ )

števila z decimalnimi mesti:

$$22,125 = 10110,001 [2] = 16,2 [16]$$

2. Iz drugih sistemov v desetiškega

Razvijemo število po sumacijski formuli (glej B.)

## D. PREDSTAVITEV NE-NUMERIČNIH INFORMACIJ V RAČUNALNIŠKEM ZAPISU

Kako lahko za računalniško rabo kodiramo (zapišemo s števili) najbolj pogosto uporabljene vrste podatkov (besedilo, zvok, glasba, film). Postopku predstavitve podatkov z dogovorjenimi znaki rečemo **kodiranje**. (Kodiranje ni prisotno samo v računalništvu, ampak pri kakršnemkoli simbolemu predstavljanju informacij – pisava, morsova abeceda, prometni znaki...).

### 1. Predstavitev (alfanumeričnih) znakov v računalniški obliki

Črke oz. znake lahko tudi prikažemo v računalniški obliki, vendar postopek ni tako enostaven kot pri številih, ki jih enostavno pretvorimo v dvojiški sistem. Alfanumerični znaki so v računalniku predstavljeni z ustreznim zaporedjem ničel in enic. Vsaka črka, vsak znak je kodirana z določeno številko (oz. zaporedjem bitov).

Koliko bitov bi potrebovali za zapis slovenske abecede?

25 znakov -> 25 različnih kombinacij ->  $\log_2 25 = 5$

□□□□□

00000 – A

00001 – B

00010 – C

00011 – Č

00100 – D

00101 – E ....

Vsako črko kodiramo oz. vsakemu znaku ustreza določeno število ali **koda**. Odločiti se je potrebno za obseg in nabor znakov, s koliko biti kodirati posamezen znak in izbranim znakom določiti dejanske kode (števila, ki jih predstavljajo). V računalništvu se je najprej oblikoval standard ASCII (*American Standard Code for Information Interchange*), pri katerem s 7 biti kodiramo 128 različnih znakov; Vsak znak je predstavljen kot kombinacija 7 znakov v dvojiškem sistemu.

Pomemben del znakov v ASCIIju so tudi t.i. kontrolne kode – znaki, ki omogočajo določene funkcije (konec vrstice; esc).

Struktura ASCII znakov:

0-31: kontrolne kode

32- 127: črke in znaki

(glej npr. <http://www2.arnes.si/~ssptvog1/ascii.htm>)

**Zakaj so standardi potrebni?** eden nujnih pogojev za komuniciranje med računalniki je enakost kodirnih tabel.

Razvoj ASCII standarda:

ASCII predpisuje 128 kod, torej je 7-bitni zapis dovolj. Ne vsebuje vseh črk slovenske abecede, zato je neposredno uporaben samo v angleško govorečih deželah. Tekom razvoja so bile ponujene določene rešitve tudi za ne-angleške znake:

- določenim kodam ASCII se je spremenil osnovni pomen (včasih se je za šumnike uporabila koda oglatih in zavutih oklepajev).
- osnovni nabor se ohrani in dopoli s prostimi kodami od 128 do 255. Pri tem je pogoj, da računalniki uporabljajo isti standard.

Na osnovi 7-bitnega ASCIIja so se razvile različne kodirne tabele, ki vključujejo jezikovne specifičnosti. Tako je danes poleg ASCIIja najbolj znana družina kodirnih tabel ISO-Latin, ki temelji na 8ih bitih.

ISO Latin 1 (tudi ISO 8859-x) je družina kodirnih tabel, ki vključuje nacionalne posebnosti - razširitev osnovnega ASCII nabora. Kodirna tabela ISO-8859-2 je namenjena za jezikovne posebnosti srednje in vzhodne evropo (in vsebuje šumnike).

Poskušajo se oblikovati tudi obsežne kodirne tabele, ki bi vključevale vse možne znake – Unicode.

Kodirni sistem operacijskega sistema Windows CP-1250, ki vključuje tudi šumnike, ni enak ISO Latin kodirni tabeli.

Današnja uporaba pojma "ascii" – ne nanaša se v tolikšni meri na star kodirni standard, temveč je oznaka za kakršnokoli datoteko, ki vsebuje čisti, surovi (plain) tekst.

## 2. Predstavitev slik (grafike) v računalniški obliki

Obstaja več načinov predstavitve slik v računalniški obliki, pri čemer so se ti načini še posebej razvili v obdobju interneta, ko je prisotna potreba po čim bolj učinkoviti predstavitvi slik (= da zasedajo čim manj bitov).

Osnovni način predstavitve slik v digitalni obliki si lahko predstavljamo kot zelo gosto mrežo majhnih kvadratkov (**pik**), kjer vsak kvadrataek predstavlja določen del slike in ima temu ustrezno barvo. V primeru črno-belih slik bodo kvadratki zgolj črni ali beli, v računalniškem jeziku to pomeni, da bodo imeli kvadratki vrednost 0 ali 1.

0	1	0
1	1	1
0	1	0

Črn križ lahko na primer predstavimo z zgornjo matriko, ki ji rečemo tudi **bitna slika** (ang. bitmap) in taka slika zavzema 9 bitov, saj za vsak kvadrataek potrebujemo le en bit. Informacija o sliki je torej predstavljena tako, da je znana informacija o vsakem elementu te matrike. Vsak tak element (oz. kvadrataek) imenujemo pika (ang.pixel).

Da je neka slika dovolj kvalitetna, mora imeti mreža dovolj kvadratkov oz. pik na določeno površino – temu rečemo tudi **ločljivost** (ang. resolution).

Sliko lahko predstavimo z različnimi barvnimi nabori (od 2, 256, 65536, do več milijonov); pri čemer za sliko z več barvami porabimo več bitov; (oz. bytov);

Če bi želeli npr. zgornji križ predstaviti z barvami iz nabora 256 barv, potem bi za celo sliko potrebovali 9 bytov oz. 72 bitov.

RGB barvni zapis. (24 bitni)

- 8 bitov za odtenek rdeče
- 8 bitov za odtenek zelene
- 8 bitov za odtenek modre

<http://www.pitt.edu/~nig/cis/web/cgi/rgb.html>

### ***Rastrska in vektorska predstavitev slike***

V računalništvu so grafični elementi predstavljeni na dva načina – slike in fotografije so ponavadi predstavljene kot matrike točk (rastrski zapis), tehnična grafika, logoti in enostavnejše slike pa s kombinacijami vektorjev (vektorski zapis). Vsak od teh dveh zapisov ima svoje prednosti in slabosti, pri čemer je ključna slabost matrične slike ravno največja prednost vektorske – pri povečevanju prvih se pokvari njihova kvaliteta, pri slednjih pa ne.

Lastnosti rastrske predstavitev slike:

- točke v matriki
- odvisna od resolucije
- pomanjševanje/povečevanje zmanjšuje kvaliteto
- omejena na pravokotnike
- kompatibilna z drugimi formati

Bitmap picture (BMP) – privzet, osnovni način rastrske slike; zelo požrešen.

JPEG in GIF – najbolj pogosta grafična formata;

JPEG - Joint Photographic Experts Group; To je tip zapisa slike, ki izjemno skompresira neko rastrsko sliko (do 96%). Uporaben za fotografije, prenos po spletu, ker je njegova velikost majhna, vendar pa na ta račun izgubimo nekaj kvalitete.

GIF – Graphical Interchange Format; GIF je bolj uporaben za ilustracije, slike z veliko količino jasne barve;



22 barvni GIF (11kb)



16 milijonov barv; JPG; 22kb

Najpogostejši programi za urejanje rastrske grafike: Adobe Photoshop; Microsoft Paint; Corel PhotoPaint; Jasc PaintShop Pro; Ulead PhotoImpact;

Lastnosti vektorske predstavitve slik:

- neodvisna od resolucije
- mogoče jo je povečevati/pomanjševati
- nima ozadja
- "risankasta"
- neprimerna za realistične slike

Najpogostejši programi za urejanje vektorske grafike: Macromedia Freehand; Adobe Illustrator, CorelDraw, Xara, Harvard draw, Autocad

### 3. Predstavitev zvoka v računalniški obliki

Zvok je izvorno analogni podatek, saj fizikalno gledano predstavlja **zvezno spreminjanje zračnega tlaka**. Graf spreminjanja tega tlaka imenujemo **zvočni val**. Računalnik neposredno analognih podatkov ne more obravnavati, zato obstajajo določeni pripomočki, ki zvok preko zvočne kartice (specifičnega elektronskega vezja) spremenijo v digitalno obliko (zvočni val ni merjen kot neprekinjen tok, ampak gre za določene zelo kratke časovne razmike). Tej pretvorbi analognega signala v digitalnega rečemo **digitalizacija**, specifična naprava, ki omogoča pretvorbo analognega zvočnega signala v digitalno obliko pa **analogno-digitalni pretvornik**.

Digitalizacija poteka tako, da računalnik v določenih presledkih (zelo kratkih) odčita velikost (amplitudo) krivulje zvočnega signala, **pogostost oz. frekvenca vzorčenja** (sampling rate - število meritev v eni časovni enoti) pa določa kvaliteto zvoka. Za zapis zvoka, ki ima CD (compact disc) kvaliteto potrebujemo 16bitni zapis s 44100 Hz vzorčne frekvence.

Digitalni zapis ni nikoli enak analognemu, vendar je na prosto uho signal praktično identičen. Izvirniku bomo bližje, če bomo opravili več meritev v eni časovni enoti.

Ker je zvok izjemno hiter fizikalni proces, je njegov računalniški zapis temu ustrezno tudi precej obsežen. In bolj kot povečujemo kvaliteto, več prostora zasede. Zato obstajajo postopki kodiranja, ki omogočajo, da originalno velikost zvočnega zapisa nekajkrat zmanjšamo (tudi do 90%).

Osnovni princip v kodiranju zvoka je dejstvo, da **nima smisla shranjevati informacije, ki jo človek tako ali tako ne more zaznati**. Instrumenti in glasbena oprema producira veliko več zvokov kot jih je človek sposoben dejansko zaznati.

#### **Dva parametra je potrebno določiti pri kodiranju:**

a) kako natančno merimo velikost zvočnega vala

b) kako pogosto izvajamo meritve;

->Maksimalne vrednosti teh dveh parametrov določa zvočna kartica – 8-bitna meri val z 256 vrednostmi, 16-bitna pa s 65.536 vrednostmi;

-> Frekvenca vzorčenja ima standardne vrednosti 11khz, 22khz in 44khz (11.000 do 44.000 meritev na sekundo).

**Kompresiranje:** Postopek rekodiranja in reduciranja zvočnega zapisa v manj obsežno obliko;

Danes najbolj znan postopek kompresiranja zvočnih informacij je **MP3 postopek**. Računalniškemu programu, ki zakodira in odkodira multimedijske podatke rečemo tudi kodek (codec - compression/decompression).

Mp3 kodek uporablja dve tehniki kompresiranja, ki omogočata redukcijo velikosti nekompresiranih digitalnih zvočnih zapisov brez, da bi veliko izgubili na kvaliteti. Mp3 odstrani tisto, kar človek ne more zaznati, poleg tega pa skompresira tudi določene "odvečnosti". Postopek je precej kompleksen: signal se razbije na posamične okvirje (frame) in ti okvirji modelirajo zvok po določenih matematičnih postopkih. Koliko kvalitete bo vseboval posamičen frame določimo sami – koliko

bitov bo rezerviranih za eno sekundo (več kot je bitov, manj bo stisnjena datoteka); Če določimo premalo bitov, bo kvaliteta zelo slaba (priporočljivo 128 – 256 kbit) Vsaka frame vsebuje podatke + glavo (metapodatki – podatki o podatkih; Poleg tega datoteka MP3 vsebuje še dodatne bite, ki vključujejo informacijo o izvajalcu ipd. (ID3).

kvaliteta zvoka	vzorčenje	način	bitov na sek.	razmerje redukcije
telefon	2.5 kHz	mono	8 kbps *	96:1
kratkovalnovni	4.5 kHz	mono	16 kbps	48:1
približno kot AM	7.5 kHz	mono	32 kbps	24:1
približno kot FM	11 kHz	stereo	56...64 kbps	26...24:1
blizu CD	15 kHz	stereo	96 kbps	16:1
CD	>15 kHz	stereo	112..128kbps	14..12:1

(Fraunhofer Institute)



#### 4. Računalniška predstavitev video (multimedijskega) zapisa

V današnjem času postajajo vedno bolj **pomembne tudi tehnologije za računalniški zapis multimedijskih informacij**, kar je posledica tega, da je vedno več dejavnosti medijev opravljenih prek interneta; po drugi strani pa šele današnja tehnologija (oz. nekaj let stara) omogoča, da se je distribucija video zapisa razširila množično med uporabnike – velikosti pomnilnika, pomnilniški mediji (CD, DVD), zapisovalniki diskov; procesorska zmogljivost (rezultat konvergence vseh teh dejavnikov).

Danes obstajajo tehnologije, ki večinoma omogočajo direktno digitalen zapis – ključni problem je, kako predstaviti ta zapis na obvladljivem prostoru in ne izgubiti preveč kvalitete.

Kodirne postopke za video je razvila skupina MPEG (Moving Picture Experts Group), pri čemer je logika kodiranja zelo podobna kodiranju zvoka – video signal je razbit na posamične "sličice" (frames), ki po matematičnih postopkih reducirajo oz. modelirajo izvorni signal; Glede na kvaliteto se odločamo o kodiranju – do približno 1.5Mbit/s. (7GB). Skupina MPEG je razvila več standardov.

MPEG-1 je najstarejši standard (.mpg datoteke); pri čemer je ena od dimenzij (audio) popularni standard za digitalno kompresijo zvoka – MP3. Mpeg-4 standard se največ uporablja v Webu; Mpeg-7 in Mpeg-32 se trenutno razvijata (večja interaktivnost; filtriranje, personalizacija; varnost, )

**DIVX** je posebni, zelo popularni način kompresiranja multimedijskih podatkov, ki uporablja standard MPEG-4. Gre za aplikacijo, ki zelo "stisne" siceršnje multimedijske datoteke in omogoča hiter prenos datotek brez velike redukcije kvalitete. (Podjetje DivXNetworks – zanimive razprave, ko je nastal ta format),.