Predstavljanje različitih tipova podataka

Cilj ove glave je da odgovori na sledeća pitanja:

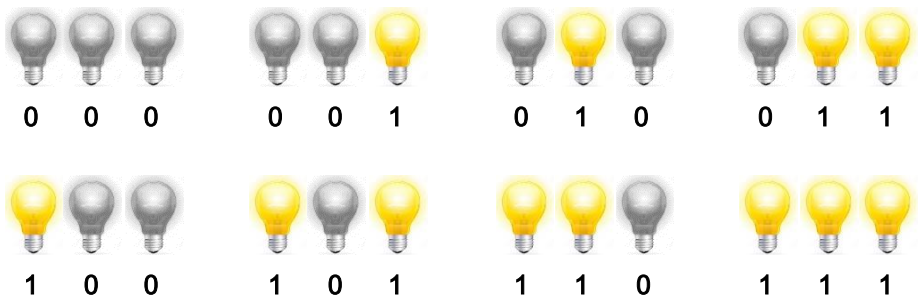
- Kako je moguće predstaviti brojeve korišćenjem isključivo nula i jedinica? Zašto je ovaj način zapisa nezaobilazan kod računara?
- Šta je digitalni, a šta analogni zapis i zašto u poslednje vreme preovlađuje digitalna tehnologija?
- Kako se u računaru zapisuju slike, zvuk i tekst? Šta su pikseli i semplovi? Šta su karakteri, a šta fontovi? Šta je ASCII, a šta Unicode?
- Kako se meri količina informacija? Šta je bit, a šta bajt? Koliko bajtova obično zauzima neki tekst, slika, pesma ili film? Koliko bajtova može da se zapiše na disku?

Računarski sistemi danas čuvaju velike količine podataka. Pitanje koje ćemo u ovoj glavi analizirati jeste kako se sve te informacije beleže u današnjim računarima. Kako se sažetak neke lektire, fotografije s ekskurzije ili film beleže u računaru? Već smo naveli da se sve ove informacije beleže u obliku brojevnih podataka, a u nastavku ćemo to detaljnije opisati.

Kodiranje podataka korišćenjem binarnog brojevnog sistema

Računari sve podatke zapisuju korišćenjem brojeva. Međutim, iz tehničkih razloga brojeve **u** memoriji računara nije moguće zapisati na uobičajeni način, pomoću deset različitih cifara, tj. nije ih moguće zapisati u dekadnom **sistemu** na koji smo navikli. Naime, memorije računara mogu se zamisliti **kao** niz elemenata koji mogu da budu **u** jednom od dva različita stanja (tzv. **bistabilni elementi).** Na primer, elektronsko kolo može biti pod naponom **ili** bez napona, deo trake može biti namagnetisan ili razmagnetisan, emulzija (metalni premaz) kompakt-diska može biti probušena ili ne. Ako stanja obeležimo različitim ciframa, moguće je koristiti samo dve različite cifre, odnosno **binarni sistem zapisa brojeva.**

Zamislimo memoriju kao niz uključenih ili isključenih sijalica. Ako zamislimo da imamo osam sijalica, da li je i kako je moguće zapisati broj 1,broj 123 ili broj 256? Ako svaku uključenu sijalicu predstavimo simbolom 1, a svaku isključenu simbolom 0 i ako pretpostavimo da svaka moguća kombinacija sijalica daje jedinstven broj, problem se svodi na pitanje da li je moguće zapisivati brojeve samo korišćenjem određenog broja nula i jedinica. Koliko različitih zapisa možemo ostvariti fiksiranim brojem cifara? Na primer, ako koristimo dve cifre, moguće je ostvariti ukupno četiri različita zapisa (to su 00, 01, 10, 11). Ako koristimo tri cifre, moguće je ostvariti ukupno osam različitih zapisa:



Može se uočiti da je pomoću k cifara moguće napraviti ukupno 2^k različitih zapisa. Kako da svakom od ovih zapisa pridružimo broj, ali tako da za dati zapis možemo lako da odredimo koji broj on predstavlja i za dati broj možemo da odredimo koji mu zapis odgovara? Da bismo dali potpun odgovor na ova pitanja, podsetićemo se uobičajenog, dekadnog, načina zapisa brojeva, a zatim ćemo taj postupak uopštiti na zapis brojeva u sistemu sa proizvoljnim brojem cifara, ili, kako se još kaže, u proizvoljnim brojevnim osnovama, a naročito u osnovi 2.

Brojevni sistemi i binarni brojevi

Već znate da se trocifreni brojevi zapisuju tako što se napiše cifra stotine, pa desetice i potom jedinice. Tako zapis 432 predstavlja broj „četiri stotine, tri desetice i dve jedinice”, tj. broj „četiristo trideset dva”.

Slično se uvode i hiljade, desetine hiljada itd. Može se primetiti da su težine svake pozicije broja u tesnoj vezi sa brojem 10. Zaista, svaka je deset puta veća od prethodne, tj. svaka predstavlja neki stepen broja 10 (na primer, 1000 je 10^3, 100 je 10^2, 10 je 10^1, a 1 je 10°). Tako je $432 = 4*10^2 + 3*10^1 + 2*10^\circ$. Prisetimo se da i različitih cifara ima tačno deset (naravno, ako računamo i nulu).

Binarni brojevi se zapisuju po potpuno istom principu, jedino što se umesto osnove 10 koristi osnova 2. U ovom slučaju postoje samo dve različite cifre, 0 i 1. Svaki broj se umesto u obliku zbira broja jedinica (10°), desetica (10^1), stotina (10^2), hiljada (10^3)... zapisuje u obliku zbira broja jedinica (2°), dvojki (2^1), četvorki (2^2), osmica (2^3) itd. Tako se broj 13 može predstaviti kao zbir jedne osmice, jedne četvorke, nula dvojaka i jedne jedinice, tj. $\mathbf{13 = 1*8 + 1*4 + 0*2 + 1*1 = 1*2^3 + 1*2^2 + 0*2^1 + 1*2^\circ}$ .Ređanjem ovih cifara dobija se zapis **1101**.

Primetimo da se osnova 10 u našem razmišljanju toliko duboko ukorenila da ni ne umemo da izgovorimo vrednost ovog broja, a da je ne razložimo na stotine, desetice i jedinice. Neko može da pomisli da je binarnih „hiljadu sto jedan" isto kao i dekadnih „trinaest", ali to je pogrešno. Ispravno je reći da se broj „trinaest" dekadno zapisuje kao **13**, a binarno kao **1101**. Da bi se nedvosmisleno znalo da je ovo binarni zapis broja „trinaest", a ne dekadni zapis broja „hiljadu sto jedan", u nekim tekstovima ovaj zapis beleži se kao $(1101)_2$ gde $(...)_2$ označava da se radi o binarnom zapisu.

Slično kao i kod dekadnog zapisa, na početak je moguće dopisivati nule, a da se vrednost broja ne promeni. Ovo možemo imati na umu ako je potrebno odrediti zapis sa unapred zadatim fiksnim brojem cifara (što je u računarima najčešće slučaj). Na primer, ukoliko se zahteva da se broj „trinaest" zapiše binarno, ali sa osam binarnih cifara, to se može uraditi ovako: $(00001101)_2$.

Za dati binarni zapis nekog broja jednostavno je odrediti koji je broj zapisan. Na primer, zapis $(10101)_2$ predstavlja zapis broja $\mathbf{1*2^4 + 0*2^3 + 1*2^2 + 0*2^1 + 1*2^\circ = 16 + 4 + 1 = 21}$.

S obzirom da se u binarnom brojnom sistemuintenzivno koriste stepeni broja 2, nije loše znati ih napamet (slično kao što se zna tablica množenja ili kvadrati brojeva do 20). Dakle, $\mathbf{2^\circ = 1, 2^1 = 2, 2^2 = 4, 2^3 = 8, 2^4 = 16 , 2^5 = 32 , 2^6 = 64 , 2^7 = 128 , 2^8 = 256 , 2^9 = 512 , 2^{10} = 1024.}$ Takođe 2^{20} je malo više od milion, 2^{30} nešto više od milijardu. Da bi se ove oznake nešto lakše izrazile, nekada se koriste oznake **K** za 2^{10}, **M** za 2^{20} i **G** za 2^{30}, pa je onda $\mathbf{2^{16} = 64K}$, $\mathbf{2^{24} = 16M}$, a $\mathbf{2^{32} = 4G.}$

Dakle, svaki prirodni broj može se na jedinstven način zapisati samo korišćenjem binarnih cifara, što onda omogućava da se podaci na jednostavan način zapišu u memorijama savremenih računara i srodnim elementima na kojima je moguće razlikovati dva stanja.

Pored osnova 10 **(dekadni zapis)** i 2 **(binarni zapis)** brojeve je J zapisivati i u drugim osnovama. Često se koriste i osnova 8 **(oktalni zapis)** i osnova 16 **(heksadekadni zapis).** Dekadni brojevi se koriste zato što imamo deset prstiju na rukama, a binarni zato što se njihov zapis tehnički jednostavno realizuje bistabilnim elementima. Oktalni i heksadekadni brojevi koriste se zato što postoje veoma jednostavni postupci da se oni prevedu u binarne brojeve i obratno, i ovi sistemi se koriste kada ljudi komuniciraju o podacima u računaru. Naime, brojevi zapisani binarno izuzetno su dugački, pa je ljudima veoma teško da rade direktno sa njima (pokušajte za vežbu da napišete broj od 32 binarne cifre, a onda samo da ga ispravno prepišete videćete da nije jednostavno). S obzirom na jednostavnost prevođenja, heksadekadni sistem obično se koristi da bi se skratio zapis dugačkih binarnih brojeva. Na primer, umesto 32 binarne cifre dovoljno je samo osam heksadekadnih. Iako ima slična svojstva, oktalni sistem se koristi malo ređe jer je stepen skraćenja zapisa malo manji.

Najvažnije brojevne osnove i njihove cifre

binarni **01**	oktalni 1234567
dekadni 0123456789	heksadekadni 0123456789ABCDEF

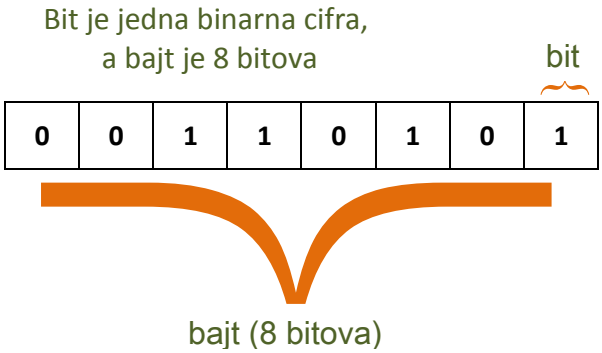
Prevođenje zapisa brojeva

Najjednostavniji način da se zapis broja prevede iz jedne u drugu osnovu jeste da se upotrebi kalkulator. Na većini računara danas, koji god operativni sistem da se koristi, unapred je instaliran neki kalkulator i većina njih podržava prevođenje zapisa prirodnih brojeva. Ovi kalkulatori obično podržavaju nekoliko režima funkcionisanja i u svakom režimu korisniku se nude različite opcije (promena režima se obično vrši menijem View ili Mode). Na primer, osnovni režim korisnicima nudi samo osnovne aritmetičke operacije, a naučni režim uključuje i elementarne matematičke funkcije (sinus, logaritam itd.). Za prevođenje zapisa brojeva obično je potrebno preći u programerski režim rada. Nakon toga, korisniku se nudi opcija da odabere neku od podržanih brojevnih osnova. Po izboru polazne osnove, ukucava se broj u toj osnovi, a na ekranu se prikazuje rezultat u ostalim podržanim osnovama. U nekim slučajevima, rezultat se ne prikazuje odmah, već korisnik nakon unosa broja bira osnovu u kojoj ga rezultat zanima i rezultat se tek tada prikazuje.

Jedinice za merenje količine podataka

Pošto se sve informacije u računaru zapisuju u obliku binarnih brojeva, osnovna jedinica mere količine podataka je broj binarnih cifara ukupno upotrebljenih za zapis. Umesto termina „binarna cifra" često se koristi skraćeni termin bit (engl. *bit – binary digit).* Na primer, ako je neki podatak zapisan pomoću 8 binarnih cifara, kažemo da on zauzima 8 bita. Bit je veoma mala količina informacija. Jednim bitom mogu se predstaviti samo broj 0 i broj 1. Već predstavljanje broja 2 zahteva dva bita (ovaj broj se zapisuje kao $(10)_2$). Da bi se kodirala bilo koja smislenija informacija (na primer, jedno slovo nekog teksta), potrebno je upotrebiti nekoliko bitova. Zato su računari tako organizovani da se za zapis svakog podatka uvek upotrebljava bar 8 bitova, odnosno jedan bajt. Memorija računara je gotovo uvek organizovana kao niz bajtova (a ne kao niz pojedinačnih bitova) i svaki bajt (a ne pojedinačni bit) ima svoj redni broj, tzv. adresu u memoriji. Stoga važi sledeća definicija:

Osnovna jedinica za merenje količine podataka je **bit.** Osnovna jedinica mere količine podataka u današnjim računarima je **bajt.**



Bajt je takođe mala jedinica jer razlikuje samo 2^8, odnosno 256 mogućnosti (i to su **$(00000000)_2$, $(00000001)_2$,... $(11111110)_2$, $(11111111)_2$).** Zato se obično u praksi koriste umnošci bajta, odnosno jedinice **kilobajt** (kB), **megabajt** (MB), **gigabajt** (GB), **terabajt** (TB).

Prefiks kilo računarstvu obično ne označava **10^3,** odnosno **1000** kao u fizici, Beć označava **2^{10},** tj. **1024.** Tako megabajt nije milion bajtova, već je nešto više od toga (odnosno **2^{20} = 1048576** bajtova). Da bi se izbegla zabuna, često se pravi razlika tako što se umesto prefiksa kilo- (k) koristi prefiks **kibi- (Ki),** umesto mega- (M) koristi se **mebi- (Mi),** umesto giga- (G) upotrebljava se **gibi- (Gi),** a umesto tera- (T) češće je **tibi- (Ti)** itd. Tako je **kibibajt (KiB)** uvek nedvosmisleno **1024 bajta.**

Nameće se pitanje koliko je bajtova potrebno da bi se zapisali neki uobičajeni sadržaji (tekst, slike, muzika, fil- movi). U nastavku ćemo detaljno opisati kako se oni zapisuju binarnim brojevima, ali i bez toga je moguće dati neke grube procene. Ako se svaki znak u tekstu predstavlja jednim bajtom (što je najčešći slučaj), onda 1 kB sadrži oko 1000 znakova. Ako jedan red teksta ima oko 80 znakova, onda 1 kB može da predstavi 25 redova čistog teksta (bez dodatnih informacija o njegovom izgledu). Boja svake tačkice slike (tzv. piksel) najčešće se zapisuje pomoću tri bajta, pa bi slika koja ima dimenziju 1000 x 1000 tačkica zauzimala oko 3 MB.

Međutim, pronađene su različite metode kompresije slike (na primer JPEG) kojima se slika zapisuje sa manje bajtova (uz neznatan gubitak kvaliteta), tako da slika ove veličine obično zauzima samo nekoliko stotina kilobajta. Slično tome, jedan minut zvuka CD kvaliteta zauzima oko 10 MB, ali kada se kompresuje može da zauzima i manje od 1 MB. Film koji traje oko dva sata, snimljen sa manjom kompresijom (kada se koristi DVD), zauzima nekoliko gigabajta, a snimljen sa većom kompresijom (obično se koristi DivX ili XviD) zauzima manje od 1 GB (stepen kompresije bira se tako da ceo film stane na kompakt-disk).

Uobičajena veličina filma, pesme, slike i teksta

Name	Size	Type
film.avi	733.4 MB	AVI video
pesma.mp3	6.0 MB	MP3 audio
slika.jpg	195.5 KB	JPEG image
tekst.txt	6.6 KB	plain text document

Uobičajeni memorijski medijumi i njihovi kapaciteti

1 TB

16 GB

4.7 GB

8 GB

Svaku memoriju karakteriše njen **kapacitet** - količina podataka koji mogu u njoj da se zapišu. Danas se ova veličina izražava u gigabajtima, pa tako diskovi računara mogu da uskladište i nekoliko stotina gigabajta, fleš-memorije nekoliko desetina gigabajta, kartice u pametnim telefonima i foto-aparatima nekoliko gigabajta, kompakt-disk oko 700 MB, DVD oko 4,7 GB itd.

Kapaciteti memorija brzo rastu s napretkom tehnologije. Na primer, kapacitet prvih fleš-memorija bio je tek nekoliko megabajta, a samo nekoliko godina kasnije njihovi kapaciteti se izražavaju u gigabajtima.

Zanimljivost

Pod pretpostaBkom da jedan neuron predstavlja jedan bit, procenjuje se da ljudski mozak skladišti i do nekoliko terabajta podataka. Naravno, reprezentacija informacija u ljudskom mozgu je drugačija nego u računarima (možemo reći da mozak efikasnije vrši skladištenje, te je količina informacija predstavljena ovom količinom podataka mnogo veća nego ona predstavljena istom količinom podataka u računaru). Takođe, postoje neke procene da lanac DNK čoveka sadrži samo nekoliko gigabajta podataka. S obzirom na to da se iste informacije zapisuju više puta, stvarna količina informacija u DNK je mnogo manja.

Digitalni zapis podataka

Analogni i digitalni časovnik

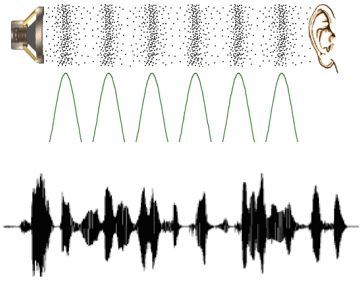
Često se kaže da živimo u **digitalnom dobu.** Većina današnjih uređaja, uključujući i sve računare, digitalna je. Koristimo digitalne foto-aparate, digitalne telefone, digitalne uređaje za reprodukciju muzike, gledamo digitalno zapisane filmove itd. Šta ovo znači? **Digitalni zapis** podrazumeva da su svi podaci zapisani u obliku brojeva. Dakle, sve što je zapisano u vašem računaru, na kompakt-disku ili fleš-memoriji zapisano je pomoću brojeva. To znači da su sve slike i pesme koje imate na računaru i svi tekstovi koje ste otkucali zapisani isključivo korišćenjem brojeva. Podaci se zapisuju na **medijumima** kao što su papir, kompakt-disk, magnetna traka itd. Danas se kao medijumi najčešće koriste različiti oblici računarskih memorija. Pre nego što su podaci počeli da se zapisuju digitalno, koristio se **analogni zapis.** Da bismo prikazali odnos analognog (kontinualnog) i digitalnog (diskretnog) zapisa, razmotrimo način snimanja muzike i način fotografisanja koje su koristili vaši roditelji u odnosu na one koje danas koristite vi. Kroz ovo poređenje objasnićemo i kako se u današnjim računarima zapisuju snimljeni zvuk i fotografije.

Digitalni zapis zvuka

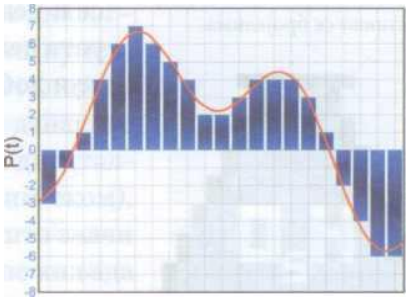
Iz fizike znate da zvuk čujemo zahvaljujući talasima koji se prostiru kroz vazduh. Zvučni talas je **kontinualan** (neprekidan) i u svakom trenutku vremena definisan je njegov intenzitet (praktično, svakom realnom broju iz nekog intervala moguće je dodeliti vrednost funkcije intenziteta zvučnog talasa). Do pre dvadesetak godina audio-snimci beležili su se isključivo na pločama ili kasetama koje sadrže magnetnu traku. Prilikom zapisa zvuka na ploču koristila se igla, koja bi vibrirala pod uticajem zvučnih talasa i u plastični materijal urezivala šaru koja im odgovara. Slično tome, glava za snimanje na kasetofonu magnetisala bi traku koja se pomerala ispred nje i količina namagnetisanja odgovarala bi intezitetu zvučnog talasa. Dakle, uspostavljana je **analogija** između fizičkog signala koji se zapisuje i onoga što je zapisano na medijumu (plastici ploče ili magnentnoj traci). Primetimo i da je ovakav zapis na medijumu kontinualan.

Ploče i kasete su se koristile za analogni zapis zvuka.

Zvuk se u vazduhu prostire kao talas. Ukoliko se predstavi grafikom, jedna pesma izgleda kao grafik na slici.



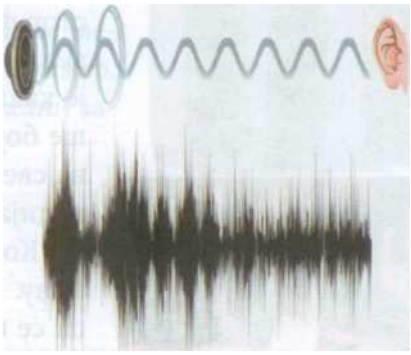
Digitalizacija analognog zvučnog signala

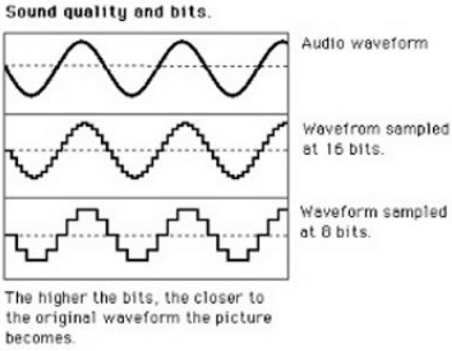


Prilikom predstavljanja zvuka u računaru nije moguće zapisati njegov intenzitet u proizvoljnom vremenskom trenutku (trenuci su kontinualni i ima ih beskonačno mnogo, a kapacitet računara je ograničen). Audio-snimci se danas zapisuju tako što uređaj za snimanje meri intenzitet zvučnog talasa u pravilnim vremenskim intervalima (i to više od 40 000 puta u sekundi) i brojeve koji iskazuju intenzitet (takozvane **semplove,** tj. **uzorke)** beleži u memoriju. Ovim dolazi do **diskretizacije** vremena (vremenski trenuci više nisu predstavljeni kontinualnim realnim vrednostima, već diskretnim decimalnim vrednostima). Međutim, i vrednost intenziteta talasa u fiksiranom vremenskom trenutku je realan broj i ne može se zapisati savršeno precizno, već se zapisuje približno. Dakle, vrši se i druga diskretizacija, ovoga puta intenziteta zvuka, a on se predstavlja korišćenjem konačnog broja dopuštenih vrednosti (za zapis se koriste dva bajta, što dopušta oko 65 000 različitih vrednosti). bolji kvalitet se postiže ako se zvuk nezavisno snima sa dva različita mikrofona i

reprodukuje nezavisno na dva zvučnika (stereo-zvuk). Sve se više koristi i višekanalno snimanje (npr. 5 + 1 sistem).

Zvuk se kroz vazduh prostire kao talas. Ukoliko se predstavi grafikom, jedna pesma izgleda kao grafik na slici.





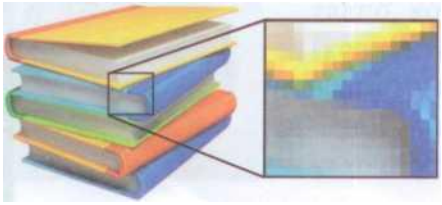
Digitalni zapis fotografija

Do pre desetak godina najčešće se koristila analogna fotografija. U aparat bi se stavljao film premazan emulzijom osetljivom na svetlost. Slika koja bi se dobila na filmu nakon njegovog razvijanja (potapanjem u različite hemikalije), doduše u negativu, odgovarala bi sceni koja se nalazila ispred foto-aparata u trenutku fotografisanja. Takođe, nakon izrade konačne fotografije (opet osvetljavanjem i hemijskim procesom), dobijena slika na papiru bi, ovoga puta direktno, odgovarala polaznoj sceni ispred foto-aparata. Postojala bi, dakle, određena **analogija** između onoga što se fotografisalo, slike na negativu i konačne slike na papiru, i ona bi bila očigledna direktnim uvidom u film ili fotografiju.

Digitalna fotografija zasnovana je na drugačijim principima. Prostor fotografije je **diskretizovan** i izdeljen na kvadratiće, koji se nazivaju **pikseli** (engl. *pixel – picture element).* Što je piksela više, slika je kvalitetnija. Zadatak foto-aparata je da za svaki piksel slike zapiše boju. Iz fizike znate da se svetlost bilo koje boje može razložiti na svetlost crvene, zelene i plave boje. Na primer, ljubičasta je kombinacija crvene i plave svetlosti, a žuta je kombinacija crvene i zelene. Kombinacija sve tri komponente daje belu, a odsustvo svetlosti daje crnu. Intenzitet svake svetlosne komponente može se zapisati brojem, pa se boja svakog piksela može predstaviti pomoću tri broja. Za zapis intenziteta svake komponente koristi se konačan broj dopuštenih vrednosti (najčešće 256 jer se za zapis koristi jedan bajt), pa je i zapis intenziteta svetlosti određene boje diskretizovan. Dakle, ako se slika posmatra kao niz piksela, a svaki piksel se predstavlja brojevima, onda se i cela slika predstavlja kao niz brojeva. Foto-ćelija digitalnog aparata deli prostor na piksele, meri količinu svake svetlosne komponente i dobijene brojeve zapisuje na memorijsku karticu foto-aparata.

Ko želi da zna više?

Za zapis fotografija kaže se da je **rasterski** *ili* **bitmapiran.** *Za zapis crteža i ilustracija obično se koristi* **vektorski zapis,** *koji podrazumeva da je slika predstavljena matematičkim (koordinatnim) opisom geometrijskih likova koji je sačinjavaju (linije, krugovi, krive)*



Digitalna slika je predstavljena sitnim tačkama - pikselima. Boja svakog piksela predstavlja se brojevima.

Definicija digitalnog i analotog zaiisa

Iz prethodna dva primera možemo da zaključimo da važi sledeća definicija:

Podaci su zapisani **digitalno** ako su zapisani pomoću niza bro**jeva.** Podaci su zapisani **analogno** ako se kontinualne promene signala zapisuju odgovarajućim kontinualnim medijuma.

Iste kasete su se koristile i za analogni zapis zvuka i za digitalni zapis programa.

Pošto se u računaru svi brojevi zapisuju binarno, često se digitalni zapis identifikuje sa binarnim zapisom. Međutim, ideja zapisa slike ili signala pomoću niza brojeva najvažniji je korak digitalizacije, dok je činjenica da se ti brojevi dalje zapisuju binarno samo tehnički detalj.

Medijum ne mora jednoznačno da određuje način zapisa, i na istim medijumima podatke je moguće zapisivati i digitalno i analogno. Na primer, magnetne trake su se koristile i za analogni zapis zvuka i za digitalni zapis računarskih programa (često video-igara).

Prednosti digitalnog zapisa podataka i poteškoće u njegovom korišćenju

Koje su prednosti digitalnog zapisa u odnosu na analogni kada je cela civilizacija proteklih decenija potpuno prešla na digitalnu tehnologiju? Ako je digitalni zapis toliko dobar, zašto se onda nije ranije koristio?

Digitalni zapis je veoma teško napraviti. Osnovni nedostatak digitalne tehnologije je to što se mnogo teže realizuje u odnosu na analognu tehnologiju. Ako se zadovoljimo lošim kvalitetom zapisa, analognu tehnologiju moguće je realizovati relativno jednostavno. vešti umetnici su portretisali i pre pojave foto-aparata, a prve fotografije, za današnje pojmove veoma lošeg kvaliteta, pojavile su se još u 19. veku. Da bi digitalna slika bila upotrebljiva (da se piskeli ne bi razlikovali golim okom), potrebno je da je sačinjava bar nekoliko stotina hiljada piksela, što znači da je pritiskom na dugme potrebno izmeriti i zabeležiti stotine hiljada brojeva. Jednostavno je uvideti da je ovakav uređaj veoma komplikovano napraviti.

Ipak, kada se prevaziđu početni tehnološki problemi, prednosti digitalnog zapisa su ogromne.

Digitalni zapis je otporan na promene medijuma. Svi materijali s vremenom se menjaju, papir žuti, a traka se razmagnetiše. Ako je analogna fotografija na papiru, kada papir požuti, pokvariće se i kvalitet fotografije. Ako je na magnentnoj traci analogno zapisan zvuk, kada se traka razmagnetiše, pokvariće se i kvalitet zvuka. Međutim, ako se na papir zapišu brojevi, kada papir požuti, informacija koju brojevi nose ostaje nepromenjena. Informacija je prisutna (u originalnom obliku) sve dok se papir toliko ne iskvari da se više ne razaznaju napisane cifre (na primer, ako ga operemo u veš-mašini). Dakle, digitalni zapis rešava problem neizbežne kvarljivosti medijuma na kojem su podaci zapisani. S obzirom na to da kvalitet zapisa nije u direktnoj vezi s kvalitetom medijuma, moguće je koristiti i nekvalitetnije i jeftinije medijume (tako je danas prosečna cena kompakt-diska znatno niža od prosečne cene kasete ranije).

Požutela i vremenom oštećena fotografija

Analogno retuširanje (levo) i alatke za digitalno retuširanje fotografija (desno)

Digitalni zapis omogućava kopiranje. Kopiranje analogno zapisanih podataka je teško. Da bi kvalitet kopije bio što verniji originalu, medijum na koji se zapisuje i uređaj kojim se kopira moraju biti veoma kvalitetni (samim tim su i skupi). Na primer, da biste iskopirali papirnu fotografiju, morate ponovo da je fotografišete, pri čemu je vrlo verovatno da će biti moguće razlikovati kopiju od originala. Kopiranje digitalno zapisanih podataka svodi se na prepisivanje brojeva. Nije neophodno da kopija bude identična originalu da bi zapisana informacija bila identična. Na primer, ako prekopirate kompakt-disk i pod mikroskopom uporedige original sa kopijom, videćete da rupice nisu apsolutno identične, ali jedino je bitno da se rupice na oba diska nalaze na istom mestu. Slično tome, ako fotokopirate svesku na kojoj pišu neki brojevi i učite te brojeve sa fotokopiranih strana, naučićete iste brojeve kao da ste učili iz originala, iako je očigledno šta je kopija, a šta je original. Dakle, digitalni zapis omogućava pravljenje kopija koje sadrže apsolutno identičnu informaciju.

Digitalni zapis omogućava jednostavnu obradu. Obrada analogno zapisanih podataka veoma je komplikovana. Da biste nešto promenili na papirnoj fotografiji, potrebno je da po njoj ili po negativu crtate četkicom i na kraju da napravite novu fotografiju. Analogna obrada zvuka podrazumeva presnimavanje zvuka korišćenjem specijalizovanih mašina (filtera), što je veoma skupo. Međutim, kada su podaci zapisani digitalno, njihova obrada podrazumeva primenu matematičkih operacija kako bi se brojevi zapisa modifikovali. S obzirom na to da računari veoma uspešno primenjuju matematičke operacije, sva obrada podataka se vrši pomoću računara i specijalizovanih programa (npr. **Adobe Photoshop i Gimp** koriste se za obradu fotografija).

Digitalni zapis omogućava jednostavan prenos. Zamislite koliko je ranije bilo komplikovano kada ste društvu želeli da pokažete kako ste se proveli na rođendanu prethodne večeri. Morali ste lično da nosite fotografije od jednog do drugog ili da napravite iste fotografije u velikom broju primeraka. Danas fotografije postavite na neku od društvenih mreža i istog trenutka celo vaše društvo može da ih vidi.

Kodiranje karaktera

U digitalnim računarima se sve, pa i tekst, predstavlja brojevima. Međutim, za razliku od slike i zvuka, koji su prirodni, kontinualni signali koji zahtevaju digitalizaciju (tj. diskretizaciju) da bi bili predstavljeni u računarima, tekst je veštačka ljudska tvorevina, koja je već diskretna, tako da ga je mnogo jednostavnije pretvoriti u niz brojeva. Tekst je u računaru uvek predstavljen kao **niz karaktera. Karakteri** su svi elementi od kojih se sastoji tekst (slova, cifre, interpunkcijski znaci i specijalni znaci, kao što su razmak ili prelazak u novi red). Da bi se tekst predstavio brojevima, vrši se **kodiranje karaktera.** Kodiranje svakom karakteru pridružuje jedinstven broj (njegov **kod).** Na primer, ako se dogovorimo da se slovo A predstavlja brojem 1, B brojem 2 itd, onda će reč BABA biti predstavljena brojevima 2 1 2 1. Da bi se tekst unet na jednom računaru i u jednom programu video na isti način u drugom programu na drugom računaru, važno je da ovaj dogovor bude univerzalan, odnosno da kodovi pridruženi karakterima budu standardizovani i da svi proizvođači računarskih programa koriste standardna kodiranja. Iz tog razloga sastavljene su standardne tablice u kojima su popisani karakteri i njihovi kodovi. Ovakve tablice nazivamo **kodne strane** ili **kodne sheme.**

U kodnoj strani, za svaki broj je naznačeno koji karakter taj broj predstavlja i kako bi taj karakter trebalo otprilike da izgleda, ali nije precizno definisan izgled karaktera. Na primer, kaže se da broju 353 odgovara malo latiničko slovo s sa kapicom (tj. slovo š), ali se ne zna da li mu odgovara slovo **Š, Š** ili š itd. Numerički kod karaktera njegova je **unutarnja reprezentacija,** dok je ono kako ga vidimo na ekranu njegova **spoljna reprezentacija.** Konkretan izgled karaktera (spoljna reprezentacija) nije određena u zapisu čistog teksta, već njega određuju programi koji čitaju zapisan tekst i prikazuju karaktere. Za ovo se obično koriste sličice koje se nazivaju **glifovi** (engl. *glyph)* i koje se prikazuju na ekranu prilikom prikaza karaktera. Pojedinačni glifovi sakupljaju se u grupe, koje se nazivaju **slovni likovi** ili **fontovi** (engl. *font*).

*Čist tekst (kreiran u najjednostavnijem programu za njegovu obradu, kao što je **Notepad**) zapisuje se prostim kodiranjem svakog njegovog karaktera. Međutim, ako je tekst formatiran (na primer, korišćenjem procesora teksta kao što je **Microsoft Word**), uz karaktere se zapisuju i dodatne informacije o izgledu teksta, kodirane na neki specifičan način, o čemu ovde neće biti reči.*

Prilikom rada sa tekstom uvek se koriste dva postupka. Dok se kuca, tekst se pretvara u niz brojeva (koristeći pritom neku kodnu stranu), a kada se prikazuje, ovaj niz brojeva pretvara se natrag u karaktere (opet uz korišćenje neke kodne strane). Sam niz brojeva ne prati uvek informacija o tome na osnovu koje kodne strane su ovi brojevi određeni, što može da dovede do problema. Može da se desi da program u kome se tekst unosi koristi jednu kodnu stranu, a da program koji tekst prikazuje koristi neku drugu. Tada se na

ekranu prikazuju karakteri koje autor nije imao na umu prilikom unosa teksta (iako su brojevi, odnosno podaci isti, njihovo tumačenje je pogrešno, pa se informacija netačno prenosi).

Verovatno ste primetili da se nekada, kada pustite film sa titlom, umesto karaktera ć prikazuje karakter ae, a umesto karaktera č prikazuje se karakter è. Radi se o tome da su ovi karakteri predstavljeni istim brojem (č i ae predstavljeni su brojem 230, a š i è brojem 232), ali u različitim kodnim stranama. Da bi se stvar popravila, potrebno je da program koji prikazuje tekst podesite tako da koristi istu kodnu stranu koju je autor koristio kada je snimao prevod filma.

Najznačajnije kodne strane

Pre određivanja brojevnih kodova za svaki pojedinačni karakter, potrebno je odrediti koji se skup karaktera uopšte kodira. Iako je poželjno da se kodira skup karaktera koji objedinjuje različita pisma (latinicu, ćirilicu, grčki alfabet, kinesko pismo itd.), nekada su računari bili skromnijeg kapaciteta i obično su dopuštali samo korišćenje manjeg skupa karaktera. S obzirom na to da su prvi elektronski računari nastali u Sjedinjenim Američkim Državama, koje su vodeća zemlja kad je računarstvo u pitanju, rani računari su dopuštali kodiranje isključivo karaktera za englesko govorno područje.

ASCII

Početkom 1960-ih, u SAD je standardizovana tablica koja se naziva **ASCII** *(American Standard Code for Information Interchange),* koja kodira 128 karaktera koji se koriste na engleskom govornom području.

Tablica ASCII osmišljena je tako da bude praktična, pa se u njoj kodovi svih cifara, svih malih i velikih slova nižu uzastopno, ili se, pak, binarni zapisi kodova odgovarajućih malih i velikih slova razlikuju tačno u jednoj cifri. Prva 32 karaktera su specijalni, kontrolni karakteri i nisu im pridruženi grafički likovi (glifovi). Većina kontrolnih karaktera danas se ne upotrebljava, ali se ipak karakteri CR i LR koriste da označe prelazak u novi red (u nekim operativnim sistemima koristi se jedan, u nekim drugi, a u nekim oba u kombinaciji). Tablica ASCII dodeljuje kodove za samo 128 različitih karaktera. Ovo znači da je dovoljno 7 bitova za zapis svakog koda, ali s obzirom na to da se računari grade tako da se podaci zapisuju kao niz bajtova, danas se za zapis svakog koda iz tablice ASCII koristi 8 bitova. Iako je donekle prevaziđena novijim standardima, tablica ASCII je predstavljala osnovu za njihov nastanak i svakako ima smisla dobro je proučiti.

*Tablica ASCII (kada se spoji broj vrste i broj kolone dobije se heksadekadni zapis koda određenog karaktera; na primer, slovo **A** ima kod $(41)_{16} = (01000001)_2 = 65$, a slovo a ima kod $(61)_{16} = (01100001)_2 = 97$).*

	LSB	0000	0001	0010	0011	0100	0101	0110	0111	1000	1001	1010	1011	1100	1101	1110	1111
MSB		0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0000	0	NUL	SOH	STX	ETX	EOT	ENQ	ACK	BEL	BS	HT	LF	VT	FF	CR	SO	SI
0001	1	DLE	DC1	DC2	DC3	DC4	NAK	SYN	ETB	CAN	EM	SUB	ESC	FS	GS	RS	US
0010	2	SP	!	"	#	$	%	&	'	(	)	*	+	,	—	.	/
0011	3	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
0100	4	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
0101	5	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	—
0110	6	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
0111	7	p	q	r	s	t	u	v	w	x	y	z	|	|	|	~	DEL

BINARY — HEX — ASCII

Unicode

Kako je kapacitet računara rastao i kako su računari počeli da se koriste u sve većem broju država, javila

se potreba za kreiranjem jedinstvenog standarda koji bi obuhvatio sve karaktere svih jezika. Danas se za ovako nešto koristi **Unicode**, nastao početkom 1990-ih. Unicode je za cilj imao da bude: univerzalan (engl. *UNIversal*) - pokriva sve savremene jezike sa pismom, jedinstven (engl. *UNIque*) - nema dupliranja karaktera jer se kodiraju pisma, a ne jezici, i uniforman (engl. *UNIform)* - svaki karakter se kodira istim brojem bitova.

Unicode u svojoj osnovnoj višejezičkoj ravni sadrži 65 536 karaktera (čiji se kodovi zapisuju pomoću dva bajta) i koji su dovoljni za zapis većine živih jezika. Prvih 128 karaktera u tablici Unicode poklapa se sa tablicom ASCII. Narednih 128 čine karakteri zapadnoevropskih latinica, nakon toga ostali latinički karakteri (među kojima su i naše Š, š, Ć, ć, Đ, đ, Č, č i Ž, ž), zatim ćirilica, grčka slova itd.

Unicode svakom karakteru dodeljuje jedinstveni numerički kod. Kodovi karaktera osnovne ravni mogu se zapisati pomoću dva bajta i ovaj način zapisa nekada se naziva **UCS-2**. Međutim, ako se na ovaj način kodira tekst koji je na engleskom jeziku i koji koristi samo ASCII karaktere, dešava se da svaki drugi bajt u zapisu ima vrednost 0 (jer ASCII karakteri i u tablici Unicode imaju veoma male brojevne vrednosti). Ovo je neracionalno i predstavlja problem starijim programima. Zato se umesto UCS-2 često koristi jedna modifikacija Unicode kodiranja koja se naziva **UTF-8** (engl. *Unicode Transformation Format).* UTF-8 za jezike koji se pišu latinicom, naročito za engleski, nudi velike prednosti u odnosu na UCS-2. Ideja je da se karakteri kodiraju različitim brojem bajtova: ASCII karakteri se kodiraju jednim bajtom (i to istim bajtom kojim bi bili kodirani i u ASCII zapisu), ostali latinički karakteri, grčka slova i ćirilički karakteri kodiraju se pomoću dva bajta, dok se, na primer, kineski karakteri kodiraju pomoću tri bajta. Postoji precizan postupak koji na osnovu broja nekog karaktera u tablici Unicode određuje bajtove kojima se taj karakter predstavlja u zapisu UTF-8.

Sažetak

U binarnom brojevnom sistemu svi brojevi se zapisuju samo ciframa 0 i 1. Računari počivaju na binarnom brojevnom sistemu zato što je njega jednostavno realizovati na bistabilnim elementima. Pored dekadnog i binarnog sistema, u računarstvu se koriste i oktalni i heksadekadni brojevni sistem zbog njihovog lakog prevođenja u binarni sistem i obratno. Konverzije iz jednog brojevnog sistema u drugi mogu se raditi kalkulatorom (koji je obično već instaliran na računaru).

Bit je jedna binarna cifra. Bajt je niz od osam binarnih cifara. Količina podataka obično se izražava u većim jedinicama od bajta (kilobajt ili kibibajt, megabajt ili mebibajt, gigabajt ili gibibajt, terabajt ili tibibajt). Tekstualni dokumenti obično zauzimaju nekoliko desetina kilobajta, slike nekoliko stotina kilobajta ili nekoliko megabajta, pesme po nekoliko megabajta, a filmovi po nekoliko stotina megabajta ili nekoliko gigabajta. Kapacitet današnjih računarskih memorija obično se izražava u gigabajtima.

Podaci su zapisani analogno ako se kontinualne promene signala zapisuju kreiranjem odgovarajućih kontinualnih promena medijuma. Podaci su zapisani digitalno kada su zapisani u vidu niza brojeva. Slika se digitalno zapisuje tako što se podeli na sitne kvadratiće (piksele), a zatim se za svaki piksel zapiše podatak o njegovoj boji. boja se zapisuje obično pomoću tri bajta, tako što se zapiše koliko crvene, zelene i plave svetlosti sadrži (RGB model). Zvuk se zapisuje digitalno tako što se njegov intenzitet meri u pravilno raspoređenim vremenskim trenucima (obično nekoliko desetina hiljada puta u sekundi) i potom se zabeleže ovi podaci (semplovi). Digitalni zapis je tehnički teško ostvariti, ali kada tehnologija dođe do tog nivoa, prednosti ovakvog zapisa su višestruke (otpornost na promene medijuma, jednostavno kopiranje, obrada, prenos).

Tekst se u računarima predstavlja kao niz karaktera (slova, cifara, interpunkcijskih znakova, ali i nekih specijalnih kontrolnih karaktera). Fontovi su zbirke glifova sličica koje se prikazuju prilikom prikaza karaktera. Čist zapis teksta sadrži samo podatke o karakterima koji ga čine, a ne i fontove. Svaki karakter se predstavlja nekim brojem, na osnovu standardizovanih tabela, koje se nazivaju kodne strane. Istorijski gledano, najznačnija tabela karaktera je ASCII sadrži samo karaktere potrebne za zapis tekstova na engleskom jeziku. Svaki karakter se kodira jednim bajtom. Danas se uglavnom koristi Unicode, koji definiše univerzalni skup karaktera (svi živi svetski jezici). Karakteri se obično kodiraju sa dva bajta (USC-2) ili sa promenljivim brojem bajtova (UTF-8).

Pitanja i zadaci za vežbu

1. Zapiši sledeće dekadne brojeve u binarnom brojevnom sistemu sa osam cifara: a) 28, b) 42, c) 123, d) 211, e) 255. Koji su brojevi određeni binarnim zapisima: a) 101101, b) 1101001, c) 1011011 ?

2. Koliko je različitih brojeva moguće zapisati pomoću četiri, osam, šesnaest binarnih cifara?

3. Zašto se u računarima svi brojevi zapisuju u binarnom brojevnom sistemu? Zašto se koristi heksadekadni sistem?

4. Šta je bit? Šta je bajt? Šta je kilobajt? Šta je kibibajt? Koliko se različitih brojeva može zapisati pomoću jednog, dva, tri i četiri bajta?

5. Koliki je kapacitet diska tvog računara ili računara u tvojoj školi? Koliki je kapacitet njegove glavne memorije? Koliki je kapacitet memorijskih kartica za digitalne foto-aparate?

6. Koliko otprilike filmova, a koliko pesama može da stane na fleš-memoriju kapaciteta 32 GB?

7. Šta znači da je neki sadržaj kompresovan? Pokušaj da saznaš koji su najčešći formati kompresovanih slika, a koji su najčešći formati kompresije muzike. Pokušaj da saznaš šta znači da je neka kompresija sa gubitkom informacije, a neka bez gubitka informacije.

8. Šta znači da je neki signal digitalno zapisan? Šta znači da je neki signal analogno zapisan?

9. Navedi i objasni osnovne prednosti digitalne tehnologije u odnosu na analognu. Zašto se digitalni zapis slika i zvuka nije koristio 1950-ih?

10. Zamisli da imaš dve flašice za mleko (sa ucrtanim podeocima). Ako u jednu uspeš 35 cl vode, a u drugu 25 cl, pa onda prespeš sve u jednu, koliko bi vode bilo u toj jednoj? Da li je ovo digitalni ili analogni uređaj za računanje? Da li su današnji računari digitalni ili analogni?

11. Šta je piksel? Šta je sempl? Koliko se bajtova obično koristi za zapis boje svakog piksela? Koji model boje se obično koristi? Koliko bajtova se obično koristi za zapis sempla?

12. Koliko piksela obično danas imaju digitalne fotografije snimljene nekim jeftinijim, a koliko one snimljene nekim skupljim foto-aparatom?

13. Šta znači da je zvuk snimljen u stereo-tehnici? Šta su sistemi 5 + 1?

14. Šta su karakteri? Šta su glifovi? Šta su fontovi? Da li je svakom karakteru pridružen glif? Kojim se karakterima obično označava prelazak u novi red?

15. Kada je nastala tablica ASCII? Koje karaktere ona definiše? Zašto računari odmah nisu koristili bogat, univerzalan skup karaktera? Kada je nastala tablica Unicode? Koje karaktere ona definiše? Zašto je uveden UTF-8? Koje su njegove prednosti u odnosu na UCS-2?

16. Kako se prema tablici ASCII kodira reč „Informatika”? Da li je moguće kodirati reč „računarstvo”? Zašto? Koliko bajtova zauzima reč „računarstvo” ako se kodira sistemom kodiranja Unicode/USC-2, a koliko sistemom kodiranja Unicode/ UTF-8? Koliko bajtova zauzima reč „računarstvo”?