

Programiranje 1

Milena Vujošević Jančić

www.matf.bg.ac.rs/~milena

Reprezentacija podataka u računaru

Pregled

- 1 Reprezentacija podataka u računaru
- 2 Zapis brojeva

Pregled

- 1 Reprezentacija podataka u računaru
 - Analogni i digitalni zapis
- 2 Zapis brojeva

Kontinualna priroda signala

- Većina podataka koje računari koriste nastaje zapisivanjem prirodnih signala.
- Signali koji nas okružuju u prirodi u većini slučajeva se prirodno mogu predstaviti neprekidnim funkcijama.
- Zvučni signal predstavlja promenu pritiska vazduha u zadatoj tački i to kao neprekidnu funkciju vremena.
- Slika se može opisati intenzitetom svetlosti određene boje (tj. određene talasne dužine) u datom vremenskom trenutku i to kao neprekidna funkcija prostora.

Analogni zapis

- Analogni zapis uspostavlja *analogiju* između signala koji je zapisan i određenog svojstva medijuma na kome je signal zapisan: kontinualne promene signala koji se zapisuje opisuju se kontinualnim promenama određenog svojstva medijuma na kojem se signal zapisuje.
- Koje analogne medijume za zapisivanje signala znate?

Analogni zapis

- Analogni zapis uspostavlja *analogiju* između signala koji je zapisan i određenog svojstva medijuma na kome je signal zapisan: kontinualne promene signala koji se zapisuje opisuju se kontinualnim promenama određenog svojstva medijuma na kojem se signal zapisuje.
- Koje analogne medijume za zapisivanje signala znate?
- Magnetna traka za zvuk i sliku, papir za sliku...

Analogni zapis

- Prednosti: ova tehnologija je obično veoma jednostavna ukoliko se zadovoljimo relativno niskim kvalitetom.
- Mane:
 - Kvalitet zapisa signala zavisi od kvaliteta medijuma — teško je napraviti kvalitetan zapis.
 - Teško je napraviti dva identicna zapisa istog signala (tj gubljenje kvaliteta prilikom pravljenja kopija) — težak je prenos signala na daljinu.
 - Primer: fotokopija fotografije je daleko lošijeg kvaliteta od originala.

Analogni zapis

- Mane:
 - Inherentna nestalnost medijuma, njegova promenljivost tokom vremena i podložnost spoljašnjim uticajima vodi do neizbežnog pada kvaliteta analogno zapisanog signala.
 - Primer: papir vremenom žuti što uzrokuje pad kvaliteta analognih fotografija tokom vremena.
 - Obrada analogno zapisanih signala je obično veoma komplikovana i za svaku vrstu obrade signala, potrebno je da postoji uređaj koji je specijalizovan za tu vrste obrade.

Digitalni zapis

- Digitalni zapis predstavlja diskretnu aproksimaciju polaznog signala.
- Vrednost signala se izmeri u određenim vremenskim trenucima ili određenim tačkama prostora i na medijumu se zapišu izmerene vrednosti.
- Ovim je svaki digitalno zapisani signal predstavljen nizom brojeva koji se nazivaju *odbirci* ili *semplovi* (*engl. sample*).
- Koliko često je potrebno vršiti merenje da bi se polazni kontinualni signal mogao verno rekonstruisati?

Digitalni zapis

- Tvrdjenje o odabiranju (tzv. Najkvist-Šenonova teorema) kaže: signal je dovoljno meriti dva puta češće od najviše frekvencije koja sa u njemu javlja.
- Postoji veliki broj prednosti digitalnog signala
- Kvalitet reprodukcije digitalnog zapisa ne zavisi od toga kakav je kvalitet medija na kome su podaci zapisani, sve dok je medijum dovoljnog kvaliteta da se zapisani brojevi mogu razaznati.

Digitalni zapis

- Kvarljivost koja je inherentna za sve medije postaje nebitna jer ne utiče na kvalitet zapisa signala, sve dok se brojevi mogu procitati kvalitet je isti.
- Primer: ako papir sadrži zapis brojeva koji predstavljaju vrednosti boja u tackama digitalno zapisane fotografije, činjenica da papir žuti ne bi predstavljala problem dok god se brojevi mogu razaznati.

Digitalni zapis

- Digitalni zapis omogućava kreiranje apsolutno identičnih kopija što dalje omogućava prenos podataka na daljinu.
- Nema gubljenja kvaliteta prilikom kopiranja: ukoliko umnožimo CD na kojem su zapisani brojevi koji čine zapis neke fotografije, kvalitet slike ostaje apsolutno isti.
- Obrada digitalno zapisanih podataka se svodi na matematičku manipulaciju brojevima i ne zahteva (za razliku od analognih podataka) korišćenje specijalizovanih mašina.

Digitalni zapis

- Osnovni problem implementacije digitalnog zapisa predstavlja činjenica da je neophodno imati veoma razvijenu tehnologiju da bi se uopšte stiglo do iole upotrebljivog zapisa.
- Na primer, izuzetno je komplikovano napraviti uređaj koji je u stanju da 40 hiljada puta izvrši merenje intenziteta zvuka.
- Jedna sekunda zvuka se predstavlja sa 40 hiljada brojeva, za čiji je zapis neophodna gotovo cela jedna sveska.
- Ovo je osnovni razlog zbog čega se digitalni zapis istorijski javio kasno.
- Kada se došlo do tehnološkog nivoa koji omogućava digitalni zapis, on je doneo mnoge prednosti u odnosu na analogni.

Pregled

1 Reprezentacija podataka u računaru

2 Zapis brojeva

- Zapis neoznačenih brojeva
- Označeni brojevi
- Zapis realnih brojeva
- Zapis teksta

Zapis brojeva

- Proces digitalizacije je proces reprezentovanja (raznovrsnih) podataka brojevima.
- Neophodno je precizno definisati zapisivanje različitih vrsta brojeva.
- U skladu sa tehnološkom osnovom digitalnih računara, osnovni sistem predstavlja *binarni* brojevni sistem (sistem sa osnovom 2).
- Koriste se i *heksadekadni* i *oktalni* brojevni sistem

Zapis brojeva

- U pozicionom brojnom sistemu, udeo cifre u celokupnoj vrednosti zapisanog broja zavisi od njene pozicije.
- Zapis broja je samo konvencija — brojevi koji se zapisuju su apsolutni i ne zavise od konkretnog zapisa.
- S obzirom na to da je svaki zapis broja u računaru ograničen, ne mogu biti zapisani svi celi i svi realni brojevi i to treba stalno imati u vidu.

Zapis neoznačenih brojeva

- Pod *neoznačenim brojevima* podrazumeva se neoznačeni zapis nenegativnih celih brojeva i znak se izostavlja iz zapisa.
- Vrednost broja zapisanog u osnovi b definiše se na sledeći način:

$$(a_n a_{n-1} \dots a_1 a_0)_b = \sum_{i=0}^n a_i \cdot b^i = a_n \cdot b^n + a_{n-1} \cdot b^{n-1} + \dots + a_1 \cdot b + a_0$$

Zapis neoznačenih brojeva

- $(9876)_{10} = 9 \cdot 10^3 + 8 \cdot 10^2 + 7 \cdot 10 + 6 = 9 \cdot 1000 + 8 \cdot 100 + 7 \cdot 10 + 6$
- $(3245)_8 = 3 \cdot 8^3 + 2 \cdot 8^2 + 4 \cdot 8 + 5 =$
 $3 \cdot 512 + 2 \cdot 64 + 4 \cdot 8 + 5 = 1536 + 128 + 32 + 5 = 1701$
- $(101101)_2 = 1 \cdot 2^5 + 0 \cdot 2^4 + 1 \cdot 2^3 + 1 \cdot 2^2 + 0 \cdot 2 + 1 =$
 $32 + 8 + 4 + 1 = 45$

Neoznačeni brojevi

- Efikasniji postupak izračunavanja se može dobiti korišćenjem *Hornerove sheme*:

$$(a_n a_{n-1} \dots a_1 a_0)_b = (\dots ((a_n \cdot b + a_{n-1}) \cdot b + a_{n-2}) \dots + a_1) \cdot b + a_0$$

- Međurezultati dobijeni u ovom računu direktno odgovaraju prefiksima zapisa čija se vrednost određuje.
- Primeri

a_i		9	8	7	6
x	0	$0 \cdot 10 + 9 = 9$	$9 \cdot 10 + 8 = 98$	$98 \cdot 10 + 7 = 987$	$987 \cdot 10 + 6 = 9876$

a_i		3	2	4	5
x	0	$0 \cdot 8 + 3 = 3$	$3 \cdot 8 + 2 = 26$	$26 \cdot 8 + 4 = 212$	$212 \cdot 8 + 5 = 1701$

Određivanje zapisa

- Za svaku cifru a_i u zapisu broja x u osnovi b važi da je $0 \leq a_i < b$.
- Pri deljenju broja x osnovom b , ostatak je a_0 a celobrojni količnik je broj čiji je zapis $(a_n a_{n-1} \dots a_1)_b$.
- Izračunavanjem celobrojnog količnika i ostatka pri deljenju sa b , određena je poslednja cifra broja x i broj koji se dobija uklanjanjem poslednje cifre iz zapisa.
- Ukoliko se isti postupak primeni na dobijeni količnik, dobija se postupak koji omogućava da se odrede sve cifre u zapisu broja x .

Određivanje zapisa

- Primer, $1701 = (3245)_8$ jer je
 $1701 = 212 \cdot 8 + 5 = (26 \cdot 8 + 4) \cdot 8 + 5 =$
 $((3 \cdot 8 + 2) \cdot 8 + 4) \cdot 8 + 5 = (((0 \cdot 8 + 3) \cdot 8 + 2) \cdot 8 + 4) \cdot 8 + 5.$

x	1701	$1701 \operatorname{div} 8 = 212$	$212 \operatorname{div} 8 = 26$	$26 \operatorname{div} 8 = 3$	$3 \operatorname{div} 8 = 0$
a_i	1701	$1701 \operatorname{mod} 8 = 5$	$212 \operatorname{mod} 8 = 4$	$26 \operatorname{mod} 8 = 2$	$3 \operatorname{mod} 8 = 3$

Direktno prevođenje binarno-heksadekadno

heksa	binarno	heksa	binarno	heksa	binarno	heksa	binarno
0	0000	4	0100	8	1000	C	1100
1	0001	5	0101	9	1001	D	1101
2	0010	6	0110	A	1010	E	1110
3	0011	7	0111	B	1011	F	1111

Na primer, proizvoljni 32-bitni sadržaj može se zapisati korišćenjem osam heksadekadnih cifara:

$$(1011\ 0110\ 0111\ 1100\ 0010\ 1001\ 1111\ 0001)_2 = (B67C29F1)_{16}$$

Zapisi fiksirane dužine

- Fiksirani broj binarnih cifara — $(\dots)_b^n$, ako se koristi n cifara.
- Ukoliko je broj cifara potrebnih za zapis broja kraći od zadate dužine zapisa, onda se broj proširuje vodećim nulama. Na primer, $55 = (0011\ 0111)_2^8$.
- Ograničavanjem broja cifara ograničava se i raspon brojeva koje je moguće zapisati

broj bitova	raspon
8	od 0 do 255
16	od 0 do 65535
32	od 0 do 4294967295

Označeni brojevi

- *Označena apsolutna vrednost i potpuni komplement.*
- Označena apsolutna vrednost: prva pozicija zapisa sadrži znak (0 označava +, 1 označava -), ostale sadrže zapis apsolutne vrednosti broja.
- Na primer, $+100 = (0\ 1100100)_2^8$, $-100 = (1\ 1100100)_2^8$.
- Problem: teško izvođenje osnovnih aritmetičkih operacija

Označeni brojevi

- Potpuni komplement:
 - Nula i pozitivni brojevi se zapisuju na isti način kao da su u pitanju neoznačeni brojevi, pri čemu u njihovom zapisu prva cifra mora da bude 0.
 - Sabiranje se sprovodi na isti način kao da su u pitanju neoznačeni brojevi, pri čemu se prenos sa poslednje pozicije zanemaruje.

	binarno	dekadno
	????????	-100
+	01100100	+100
(1)	00000000	0

Označeni brojevi

- $x + (-x) = (1\ 00 \dots 00)_2^{n+1}$
- Pošto je $(1\ 00 \dots 00)_2^{n+1} = (11 \dots 11)_2^n + 1$, zapis broja $(-x)$ je moguće odrediti tako što se izračuna $(11 \dots 11)_2^n - x + 1$. Izračunavanje razlike $(11 \dots 11)_2^n - x$ se svodi na *komplementiranje* svake pojedinačne cifre broja x .

01100100	+100
10011011	komplementiranje
+ 1	
10011100	-100

- Ne važi samo za zapis broja $(100 \dots 00)_2^n$ koji je sam sebi komplementaran.

Označeni brojevi — raspon

broj bita	raspon
8	od -128 do $+127$
16	od -32768 do $+32767$
32	od -2147483648 do $+2147483647$

Zapis realnih brojeva

- Komplikovanije je predstaviti relane brojeve nego cele brojeve
- Moguće je predstaviti samo određeni podskup skupa realnih brojeva i to podskup skupa racionalnih brojeva.
- Interni zapis celog broja i njemu odgovarajućeg realnog se ne poklapaju (nule i jedinice kojima se oni zapisuju nisu iste), čak i kada se isti broj bitova koristi za njihov zapis.
- U zapisu realnih brojeva je potrebno napraviti kompromis između širine raspona brojeva koji se mogu zapisati (slično kao i kod celih brojeva), ali i između preciznosti brojeva koji se mogu zapisati.

Zapis realnih brojeva

- Osnovni načini zapisivanja realnih brojeva su zapis u *fiksnom zarezu* i zapis u *pokretnom zarezu*.
- Zapis u fiksnom zrezu podrazumeva da se posebno zapisuje znak broja, zatim ceo deo broja i zatim njegov razlomljeni deo (njegove decimale).
- Broj cifara za zapis celog dela i za zapis realnog dela je fiksiran i jednak je za sve brojeve u okviru istog tipa koji koristi zapis u fiksnom zrezu.

Zapis realnih brojeva

- Zapis u pokretnom zarezu podrazumeva oblik $\pm m \cdot b^e$.
- b je osnova (2), m se naziva mantisa, a vrednost e naziva se eksponent.
- Broj cifara (obično binarnih) za zapis mantise i broj cifara za zapis eksponenta je fiksiran i jednak je za sve brojeve u okviru istog tipa koji koristi zapis u pokretnom zarezu.
- Standard *IEEE 754* — predviđa da se određene kombinacije bitova odvoje za zapis tzv. specijalnih vrednosti (beskonačnih vrednosti, grešaka u izračunavanju i slično).

Zapis realnih brojeva

- Fiksni zarez ima jednaku preciznost na celom rasponu
- Porektni zarez ima veću gustinu predstavljenih malih brojeva nego kod velikih

Zapis teksta

- Tekst: „informaciju namenjenu ljudskom sporazumevanju koja može biti prikazana u dvodimenzionalnom obliku.”
- u računarima se tekst predstavlja kao jednodimenzioni (linearni) niz karaktera koji pripadaju određenom unapred fiksiranom skupu karaktera.
- U zapisu teksta, koriste se specijalni karakteri koji označavaju prelazak u novi red, tabulator, kraj teksta i slično.
- Osnovna ideja koja omogućava zapis teksta u računarima je da se svakom karakteru pridruži određen (neoznačeni) ceo broj i to na unapred dogovoreni način.
- Ovi brojevi se nazivaju *kôdovima karaktera*

Zapis teksta

- Postoji više različitih standardnih tabela koje dodeljuju numeričke kôdove karakterima.
- Postoje 7-bitni kôdovi, 8-bitni kôdovi, 16-bitni kôdovi, 32-bitni kôdovi, kao i kodiranja promenljive dužine koja različitim karakterima dodeljuju kôdove različite dužine
- Tabele koje sadrže karaktere i njima pridružene kôdove obično se nazivaju *kôdne strane*
- Najvažniji standardi: ASCII, UNICODE, UCS2, UTF8

Zapis teksta

- Osobine ASCII standarda — 7 bita
 - Prva 32 karaktera – od $(00)_{16}$ do $(1F)_{16}$ – su specijalni kontrolni karakteri.
 - Ukupno 95 karaktera ima pridružene grafičke likove (engl. printable characters).
 - Cifre 0-9 predstavljene su kôdovima $(30)_{16}$ do $(39)_{16}$, tako da se njihov ASCII zapis jednostavno dobija dodavanjem prefiksa 011 na njihov binarni zapis.
 - Kôdovi velikih i malih slova se razlikuju u samo jednom bitu u binarnoj reprezentaciji. Na primer, *A* se kodira brojem $(41)_{16}$ odnosno $(100\ 0001)_2$, dok se *a* kodira brojem $(61)_{16}$ odnosno $(110\ 0001)_2$.
 - Slova su poredana u *kolacionu sekvencu*, u skladu sa engleskim alfabetom.

Zapis teksta

- Postoje 8-bitna proširenja ASCII standarda, od strane ISO ali i od Microsoft-a
- Iako 8-bitne kôdne strane omogućavaju kodiranje tekstova koji nisu na engleskom jeziku, nije moguće, na primer, u istom tekstu koristiti i ćirilicu i latinicu.
- Takođe, za azijske jezike nije dovoljno 256 mesta za zapis svih karaktera

Zapis teksta

- Standardi ISO 10646 i UNICODE — 4 bajta za karakter
- UCS2 — preslikava deo UNICODE karaktera u dva bajta
- UTF8 — je algoritam koji svakom dvobajtnom Unicode karakteru dodeljuje određeni niz bajtova čija dužina varira od 1 do najviše 3. UTF je ASCII kompatibilan, što znači da se ASCII karakteri zapisuju pomoću jednog bajta, na standardni način.

Literatura

Slajdovi su pripremljeni na osnovu materijala iz drugog poglavlja knjige:

Predrag Janičić, Filip Marić: Programiranje 1