

3. Podaci i informacije

Obrada podataka

Pod obradom podataka podrazumeva se proces kojim se podaci transformišu tako da se od nekog početnog skupa (ulaznih) podataka dobije novi (izlazni) skup podataka. Tako se recimo mogu sabrati visine (ili težine) niza osoba, pa kada se takav zbir podeli sa brojem osoba čije su visine (težine) sabrane dobijamo informaciju o prosečnoj visini (težini) tog skupa osoba. Može se reći da smo obradom podataka o visinama (težinama) kao rezultat dobili jednu novu informaciju (podatak). Zapravo svaka obrada podataka i ima za cilj dobijanje nove informacije koja je “sakrivena” u sirovim podacima.

Proces obrade podataka možemo slikovito prikazati sledećim dijagramom:



Slika 3.1 Proces obrade podataka

Podrazumeva se, naravno, da se obrada podataka vrši u kompjuterima. Podaci se mogu u kompjuterima obrađivati na više načina u zavisnosti od uslova pod kojima se obrada vrši.

Ako se obrada vrši na taj način što se podaci najpre prikupljaju, prilagođavaju za kompjutersku obradu unose u kompjuter tek kada su svi podaci prikupljeni onda takvu obradu podataka nazivamo **batch** (beč) obradom. To je slučaj sa raznim statističkim obradama, kao i obradama recimo plata u nekom preduzeću, obradama računa za električnu energiju, TV pretplatu i slično. Trajanje ovakve obrade može biti i više sati rada kompjutera.

Ako se obrada vrši odmah nakon unosa podataka onda kažemo da se radi o **on-line** obradi. To je slučaj recimo sa bankarskim transakcijama kada podižete (ili ulažete) gotovinu sa vašeg računa, plaćate kreditnom karticom ili rezervišete mesto u avionu (ili pozorištu). Kod takvi obrada se ne može čekati na prispeće svih podataka, već se obrada vrši redom kako oni stižu. Trajanje obrade može biti do nekoliko sekundi (ponekad i minuta).

Postoje slučajevi kada se podaci moraju obraditi za veoma kratko vreme (ispod 1 sekunde ili čak za hiljaditi deo sekunde). Za takve obrade kažemo da se odvijaju u realnom vremenu (**real-time**). Primeri takve obrade su razne vrste upravljanja robotima, automatskim pilotima, hirurškim zahvatima i sl. Te su obrade često povezane i sa visokim rizicima po bezbednost ljudi ili gubitak njihove imovine.

A kakvi se sve podaci mogu obrađivati? Kakve sve oblike podataka znamo?

Vrste podataka i način kodiranja

Podaci i informacije imaju dve važne karakteristike: formu i sadržinu. Forma je oblik u kojem se podaci pojavljuju a sadržina je njihovo značenje. Uobičajene i najrasprostranjenije forme podataka su slova, brojevi, zvuk, slika, filmski zapis. To su forme prepoznatljive čoveku. Čovek može da prepozna i neke druge forme podataka koje prima i drugim čulima kao što su dodir, miris, ukus, temperatura. Ali u prirodi postoje i druge fizičke karakteristike materije i prostora koje su moguće forme podataka kao što su električni i magnetni signali, bio-energetski signali da pomenemo samo neke od njih.

Naša dalja pažnja biće usmerena samo na standardne forme podataka koje su najčešće u upotrebi, mada se taj skup stalno širi novim formama.

Značenje podataka predstavlja drugu važnu njihovu karakteristiku. Jedan te isti podatak može imati različita značenja, u zavisnosti od interpretacije. Kakvu ćemo interpretaciju pridružiti nekom podatku? Najčešće je interpretacija povezana sa kontekstom u kojem se neki podatak nalazi. Mi se ovde nećemo baviti značenjima podataka već samo njihovom formom i mogućnostima transformacije podataka iz jedne forme u drugu, kao i kombinovanjem podataka u cilju dobijanja novih informacija onako kako to prikazuje gornji dijagram.

Numerički podaci (brojevi)

Brojevima se izražavaju kvantifikatorska svojstva objekata i procesa. Brojevima prikazujemo cene, količine, fizičke veličine (visine, težine, temperature, itd.) kao i razne druge merljive podatke. Brojevima se mogu modelirati i veoma složene prirodne i društvene pojave i procesi. Statistički pokazatelji su dobar primer za upotrebu brojeva pri analizi prirodnih i društvenih fenomena. Aritmetika je oblast matematike koja se bavi osnovnim karakteristikama brojeva. Tokom razvoja matematike uvedeno je više vrsta brojeva, od prirodnih brojeva (0,1,2,3...), celih brojeva (...-2,-1,0,1,2,...), racionalnih brojeva (razlomaka, $\frac{1}{2}$, $\frac{2}{3}$, ...), pa sve do kompleksnih brojeva ($2 + 3i$). Za nas su najinteresantniji celi i decimalni brojevi kao brojevi koji se najčešće koriste u praksi.

Dekadni brojni sistem

Brojevi se mogu zapisivati na više načina. U svakodnevnoj upotrebi je tzv. dekadni brojni sistem gde se za zapisivanje brojeva koristi deset cifara (0,1,2,3,4,5,6,7,8,9), pa otuda i ime dekadni (deka na grčkom znači deset).

U dekadnom sistemu se recimo zapis 327 tumači kao broj: tri stotine dvadest sedam. To se može i ovako prikazati:

$327 = 3 \times 100 + 2 \times 10 + 7$ (tri puta sto plus 2 puta deset plus 7), ili ovako:

$327 = 3 \times 10^2 + 2 \times 10^1 + 7 \times 10^0$ (tri puta deset na kvadrat plus dva puta deset na jedan plus 7 puta deset na nula), pa zato kažemo da je ovo pozicioni brojni sistem jer vrednost cifre zavisi od mesta na kojem se cifra nalazi u broju.

Decimalni brojevi pored cifara koriste i decimalni zarez kojim se razdava broj celi od broja koji označava deo celog. Tako recimo broj 1835,346 označava broj koji ima 1835 celih i dodatak od 346 hiljaditih delova celog (346/1000). U kompjuterskim aplikacijama decimalni zarez se često zamenjuje decimalnom tačkom tako da se gornji broj može zapisati i kao 1835.346.

Binarni brojni sistem

Brojevi se na sličan način mogu zapisivati sa različitim brojem cifara. U kompjuterskim naukama u upotrebi su pored dekadnog još i binarni, oktalni i heksadecimalni brojni sistemi.

Binarni brojni sistem koristi samo dve cifre: 0 i 1. Ovaj sistem je pogodan za komjutere zbog tehnologije njihove izrade i tzv. Bulove algebre na kojoj se zasniva kompjuterska tehnika. Slično kao i kod dekadnog sistema brojevi se zapisuju ciframa 0 i 1, a vrednost cifre zavisi od njene pozicije u broju. Tako na primer, binarni broj 10011 zapravo ima sledeću vrednost:

$$10011 = 1x2^4 + 0x2^3 + 0x2^2 + 1x2^1 + 1x2^0$$

Ili u dekadnom obliku $1x16 + 0x8 + 0x4 + 1x2 + 1x1 = 19$ (devetnaest).

Gornji primer pokazuje kako se vrši konverzija iz binarnog u dekadni brojni sistem. Potrebno je, zanači, samo svakoj cifri binarnog broja pridružiti njenu vrednost u zavisnosti od pozicije i sabrati.

Obrnute konverzija, iz dekadnog u binarni oblik se može izvršiti sledećim postupkom:

1. Broj koji se konvertuje podeliti sa 2.
2. Rezultat deljenja zapisati ispod broja koji se konvertuje.
3. Ostatak pri deljenju (koji može biti samo 0 ili 1) zapisati sa strane.
4. Ponoviti korake 1-3 sve dog rezultat deljenja ne bude jednak 0.
5. Ostatke pri deljenju (nule i jedinice) poredjati jedne za drugim u redosledu od poslednjeg do prvog izračunatog ostatka.

Gornji postupa ilustrujmo na primeru konverzije broja 213 i binarni oblik.

213	podeljeno sa 2	daje rezultat	106	i ostatak	1
106	podeljeno sa 2	daje rezultat	53	i ostatak	0
53	podeljeno sa 2	daje rezultat	26	i ostatak	1
26	podeljeno sa 2	daje rezultat	13	i ostatak	0
13	podeljeno sa 2	daje rezultat	6	i ostatak	1
6	podeljeno sa 2	daje rezultat	3	i ostatak	0
3	podeljeno sa 2	daje rezultat	1	i ostatak	1
1	podeljeno sa 2	daje rezultat	0	i ostatak	1

Dakle, 213 dekadno je isto što i 11010101 binarno.

Heksadecimalni brojni sistem

Heksadecimalni brojni sistem ima 16 cifara (heksa na grčkom znači šestnaest).

Pored deset cifara koje su preuzete iz dekadnog sistema heksadecimalni sistem koristi i slova A,B,C,D,E i F. To je takođe pozicionin brojnin sistem sa ciframa: 0,1,2,3,4,5,6,7,8,9,A,B,C,D,E,F, gde cifre 0-9 imaju uobičajene vrednosti a A,B,C,D,E i F vrednosti 10,11,12,13,14, i 15 redom.

Tako je, na primer, heksadecimalni broj 4A2 zapravo:

$$4A2 = 4 \times 16^2 + 10 \times 16^1 + 2 \times 16^0 \quad (\text{setiete se da A ima vrednost 10, otuda } 10 \times 16^1)$$

Dakle 4A2 heksadecimalno ima vrednost 1186 (1024 + 160 + 2).

Konverzija iz dekadnog u heksadecimalni oblik (i obrnuto) vrši se analogno kao kod binarnih brojeva.

Međutim, konverzija iz binarnog u heksadecimalni oblik (i obrnuto) je mnogo jednostavnija.

Recimo da želimo da heksadecimalni broj 4A2 (iz gornjrg primera) prevedemo u binarni oblik. Sve što treba da uradimo je da svaku heksadecimalnu cifu prevedemo u binarni oblik, kako sledi:

4A2 hesadecimalno je 100101010 binarno jer je 4 u binarnom obliku 100, A je 1010, a 2 je 10, pa kad ova tri binarna niza spojimo dobijamo 100101010.

Obrnuto, ako želimo da neki binarni broj prevedemo u hesadecimalni oblik, sve što treba da uradimo je da binarni broj razdelimo u grupe od po četiri cifre (počev od krajnje desne cifre) i svaku takvu grupu prevedemo u heksadecimalnu cifru.

Na primer binarni broj 1000101010101110100100111 koji na prvi pogled izgleda zastrašujuće dugačak, lao se prevodi u heksadecimalni zapis.

Najpre broj razdelimo u grupe od po četiri cifre. Dobićemo sledeći niz:

1 0001 0101 0101 1101 0010 0111

Sada svaku grupu prevedimo u heksadecimalnu cifu:

Iamaćemo:

1 prevodimo u 1
0001 takođe u 1
0101 prevodimo u 5
0101 takođe u 5
1101 prevodimo u D (13)
0010 prevodimo u 2
0111 prevodimo u 7

Pa je tako

1000101010101110100100111 binarno jednako 1155D27 heksadecimalno. Lako, zar ne?

Operacije sa brojevima

Brojeve možemo da sabiramo, oduzimamo, delimo i možimo. To su tzv. osnovne aritmetičke operacije sa brojevima. Već se u početnom stadijumu učenja srećemo sa ovim operacijama i naučimo kako da ih izvršavamo nad dekadno zapisanim brojevima. Ovde ćemo samo napomenuti da se gornje aritmetičke operacije izvršavaju analognim postupcima i za binarne i heksadecimalno zapisane brojeve.

I najsloženiji matematički proračuni, najsloženije jednačine se na kraju svode na ove četiri operacije. Sve što se može sračunati sračunava se ovim operacijama. Naravno za složene proračune imamo postupke (algoritme) u kojima se ove osnovne operacije kombinuju na različite načine da se dođe do željenih rezultata. Ali o algoritmima ćemo kasnije. O brojevima i operacijama sa brojevima za sada ovoliko.

Tekstualni podaci

Naša civilizacija je dobila ubrzan razvoj od trenutka kada je čovek počeo da zapisuje svoje misli u obliku teksta. Posebno ubrzanje je dobijeno Gutenbergovim pronalaskom štamparije kada je postalo moguće tekst umnožavati u više primeraka na ekonomičan način. Pojavom kompjutera, a posebno personalnih kompjutera i njihovim povezivanjem u globalnu mrežu (Internet) dostupnost tekstualnih informacija u obliku knjiga, časopisa i drugih tekstova postala je tako velika da su nam danas potrebni posebni pretraživači (Google na primer) bez čije pomoći bi bili izgubljeni u tom okeanu informacija, među kojima tekstualne informacije imaju značajan udeo.

Pa kako su računari u stanju da pamte tekst. Za brojeve smo već videli, koriste cifre 0 i 1. A tekst? Takođe. Kako je to moguće?

Na samom početku korišćenja računara (50-tih godina prošlog veka) napravljen je jedan šifarnik kojim su slova kodirana brojevima. Pa kada slova prevedemo u brojeve onda je lako, brojeve možemo zapisivati u kompjuterima.

Napravljeni šifarnik slova poznat je kao ASCII (American Standard Code for Information Interchange) tabela i još uvek je u upotrebi (vidi sledeću tabelu).

Dec	Hx	Oct	Chr		Dec	Hx	Oct	Html	Chr	Dec	Hx	Oct	Html	Chr	Dec	Hx	Oct	Html	Chr
0	0	000	NUL	(null)	32	20	040	##32;	Space	64	40	100	##64;	@	96	60	140	##96;	`
1	1	001	SOH	(start of heading)	33	21	041	##33;	!	65	41	101	##65;	A	97	61	141	##97;	a
2	2	002	STX	(start of text)	34	22	042	##34;	"	66	42	102	##66;	B	98	62	142	##98;	b
3	3	003	ETX	(end of text)	35	23	043	##35;	#	67	43	103	##67;	C	99	63	143	##99;	c
4	4	004	EOT	(end of transmission)	36	24	044	##36;	\$	68	44	104	##68;	D	100	64	144	##100;	d
5	5	005	ENQ	(enquiry)	37	25	045	##37;	%	69	45	105	##69;	E	101	65	145	##101;	e
6	6	006	ACK	(acknowledge)	38	26	046	##38;	&	70	46	106	##70;	F	102	66	146	##102;	f
7	7	007	BEL	(bell)	39	27	047	##39;	'	71	47	107	##71;	G	103	67	147	##103;	g
8	8	010	BS	(backspace)	40	28	050	##40;	(72	48	110	##72;	H	104	68	150	##104;	h
9	9	011	TAB	(horizontal tab)	41	29	051	##41;)	73	49	111	##73;	I	105	69	151	##105;	i
10	A	012	LF	(NL line feed, new line)	42	2A	052	##42;	*	74	4A	112	##74;	J	106	6A	152	##106;	j
11	B	013	VT	(vertical tab)	43	2B	053	##43;	+	75	4B	113	##75;	K	107	6B	153	##107;	k
12	C	014	FF	(NP form feed, new page)	44	2C	054	##44;	,	76	4C	114	##76;	L	108	6C	154	##108;	l
13	D	015	CR	(carriage return)	45	2D	055	##45;	-	77	4D	115	##77;	M	109	6D	155	##109;	m
14	E	016	SO	(shift out)	46	2E	056	##46;	.	78	4E	116	##78;	N	110	6E	156	##110;	n
15	F	017	SI	(shift in)	47	2F	057	##47;	/	79	4F	117	##79;	O	111	6F	157	##111;	o
16	10	020	DLE	(data link escape)	48	30	060	##48;	0	80	50	120	##80;	P	112	70	160	##112;	p
17	11	021	DC1	(device control 1)	49	31	061	##49;	1	81	51	121	##81;	Q	113	71	161	##113;	q
18	12	022	DC2	(device control 2)	50	32	062	##50;	2	82	52	122	##82;	R	114	72	162	##114;	r
19	13	023	DC3	(device control 3)	51	33	063	##51;	3	83	53	123	##83;	S	115	73	163	##115;	s
20	14	024	DC4	(device control 4)	52	34	064	##52;	4	84	54	124	##84;	T	116	74	164	##116;	t
21	15	025	NAK	(negative acknowledge)	53	35	065	##53;	5	85	55	125	##85;	U	117	75	165	##117;	u
22	16	026	SYN	(synchronous idle)	54	36	066	##54;	6	86	56	126	##86;	V	118	76	166	##118;	v
23	17	027	ETB	(end of trans. block)	55	37	067	##55;	7	87	57	127	##87;	W	119	77	167	##119;	w
24	18	030	CAN	(cancel)	56	38	070	##56;	8	88	58	130	##88;	X	120	78	170	##120;	x
25	19	031	EM	(end of medium)	57	39	071	##57;	9	89	59	131	##89;	Y	121	79	171	##121;	y
26	1A	032	SUB	(substitute)	58	3A	072	##58;	:	90	5A	132	##90;	Z	122	7A	172	##122;	z
27	1B	033	ESC	(escape)	59	3B	073	##59;	;	91	5B	133	##91;	[123	7B	173	##123;	{
28	1C	034	FS	(file separator)	60	3C	074	##60;	<	92	5C	134	##92;	\	124	7C	174	##124;	
29	1D	035	GS	(group separator)	61	3D	075	##61;	=	93	5D	135	##93;]	125	7D	175	##125;	}
30	1E	036	RS	(record separator)	62	3E	076	##62;	>	94	5E	136	##94;	^	126	7E	176	##126;	~
31	1F	037	US	(unit separator)	63	3F	077	##63;	?	95	5F	137	##95;	_	127	7F	177	##127;	DEL

Source: www.LookupTables.com

Slika 3.2 ASCII tabela

Kao što se može videti sa gornje slike ASCII tabela sadrži slova internacionalnih alfabeta (26 slova od A do Z, mala i velika), cifre znake interpunkcije I razne druge simbole, sa ili bez grafičkog prikaza.

Tako recimo veliko slovo A ima šifru 65 u dekadnom brojnom sistemu, ili 41 u heksadecimalnom.

ASCII kod je prvobitno bio 7-bitni kod (jer ima 128 znakova, a oni se mogu kodirati sa 7 binarnih cifara). Pošto je sadržao samo internacionalnu latiničnu azbuku, ovaj kod je bilo nemoguće koristiti za druge jezike i pisma (osim engleskog).

U poslednje vreme, a posebno za potrebe interneta koristi se novi tzv. UNICODE. Ovaj kod se pojavljuje u više varijanti, a UTF-16 je 16-bitna varijanta koja pokriva 30-tak sistema za pisanje teksta na raynim jezicima.

Operacije sa tekstom

Slično kao što kod brojeva postoje osnovne operacije, tako i obrada tesktualnih podataka ima jedan skup jednostavnih operacija pomoću kojih se mogu obavljati i najsloženije operacije nad tesktovima. To su operacije nad pojedinačnim slovima u tekstu kao što su: brisanje slova, dodavanje slova, promena slova. Pomoću tih operacija moguće je izvršiti razne operacije nad celim tekstovima kao što su brisanje dela teksta, dodavanje teksta, zamena teksta drugim tekstom. Tekst procesori (MS Word, na primer) obavljaju slične operacije.

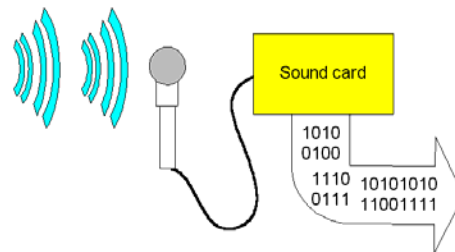
Razvojem grafike (kako na ekranima tako kod štampača) postalo je moguće prikazivati slova u raznim oblicima (fontovima), raznim veličinama i drigim tipografskim karakteristikama (boja, otisak, italik i sl.).

Danas se u kompjuterima koriste i tzv. hipertekstovi u kojima se tekst ne mora čitati lierno već se može “skakati” sa teksta na tekst u proizvoljnom redosledu sledeći takoznake hiperlinkove. A tekstualni zapisi mogu da sadrže slike, video i zvučne delove, tako da se dobija informacija obogaćena raznim vrstama i formama podataka.

Zvučni podaci

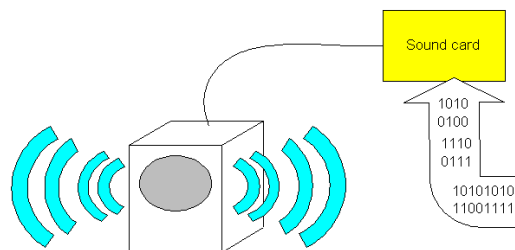
Zvuk, kao što je poznato nastaje vibracijom vazduha, pa kada se ta vibracija prenese do naše bubne opne, ona vibrira u istom ritmu i proizvodi signal koji naš mozak interpretira kao zvuk.

Kako zvuk registruju, pamte i reprodukuju kompjuteri? Znamo da su mikrofoni i zvučnici standardni deo periferije naših PC-ja. No kako se zvuk pamti u kompjuteru? Opet pomoću brojeva. Posto je zvuk vibracija vazduha, ta se vibracija u mikrofону (slično kao u našoj bubnoj opni) pretvara u električni signal kako je to ilustrovano na sledećoj slici.



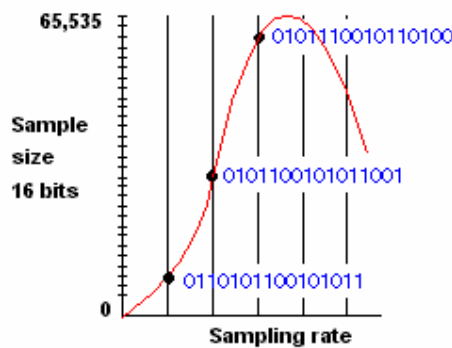
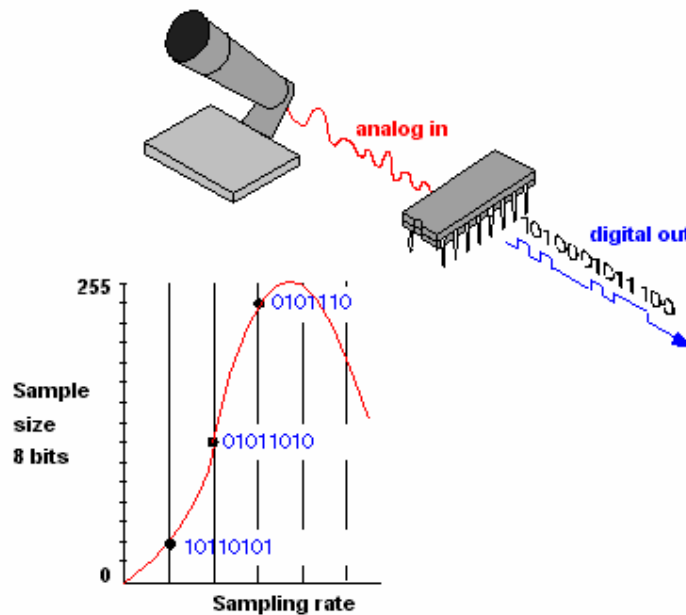
Slika 3.3 Registrovanje zvuka

S druge strane, reprodukcija zvučnog signala se odvija tako što se brojevi kojima je registrovan zvuk pretvaraju ponovo u električna signale kako to prikazuje sledeća slika.



Slika 3.4 Reprodukija zvuka

Konverzija iz zvuka u brojeve (i obrnuto) vrši se tako što se analogni električni signali sempliraju (digitalizuju) na način ilustrovan sledećom slikom.



3.5 Digitalizacija zvuka

Zvučni podaci (govor, muzika) se mogu takođe obrađivati. I kod njih postoji jedan skup osnovnih operacija: brisanje dela zvučnog zapisa, dodavanje zvučnog zapisa, zamena zvučnog zapisa. Pomoću takvih operacija moguće je izvršiti veoma složene transformacije zvučnih zapisa i postići razne zvučne efekte.

U praksi se koristi više načina digitalizacije i zapisivanja zvučnih podataka. Kompjuterski fajlovi u kojima se nalazi zvuk mogu imati razne ekstenzije kao što su wav, mp3, aud, midi, da pomenemo samo neke od najpoznatijih.

Grafika (slika, video)

Jedna slika vredi hiljadu reči kaže kineska poslovice. Vizuuelne informacije su veoma pogodne za čoveka. Zato kompjuterska tehnologija posebnu pažnju poklanja ovoj vrsti podataka. Svedoci smo stalnog razvoja u ovoj oblasti, gde postoji tendencija da

Slično, ako umesto kolora želimo sliku sa tonovima sivog možemo svakom pikselu pridružiti jedan bajt kojim se opisuje jedan od 256 nivoa sivog.

A šta ako imamo pokretnu sliku (video zapis). Pa nije problem, potrebno je samo pamtiti više uzastopnih slika u jednoj sekundi. Kod klasičnih filmova sa filmskom vrpcom uobičajeno je da se prave 24 slike u sekundi, kod kompjutera taj broj je obično 36 pa je tako kompjuterska slika stabilnija.

Naravno zapisivanje slike, a posebno videa zahteva veliku količinu kompjuterske memorije, mnogo veće nego za zvuk, a naročito u odnosu na tekst i brojeve.

I sliku (video) je moguće obrađivati. Osnovne operacije su ovde promena karakteristika piksela (boje, osvetljaja, kontrasta).

Postoji takođe, mnogo načina digitalizacije slike (i vide-a). Najpoznatiji formati zapisa su bmp, jpg, png, pic, tiff itd.

Pitanja

1. Kako se podaci razlikuju od informacija?
2. Nevedite osnovne operacije koje su prisutne u informacionim sistemima?
3. Od kojih se komponenti sastoji informacioni sistem? Šta sve IS uključuje?
4. Navedite neke od operacija koje se mogu vršiti nad podacima?
5. Navedite bar tri vrste podataka.
6. Navedite najčešće korišćene osnovne tipove podataka u programskim jezicima.
7. Definišite podatke tipa bool(ean).
8. Koliko je heksadecimalno 7F u dekadnom sistemu?
9. Koliko bajtova je potrebno za predstavljanje jednog slova u ASCII kodu? A u Unicode-u.
10. Zašto služi kompresija zvučnih signala?
11. Navedite bar dva formata zvučnog zapisa.
12. Navedite bar tri formata za memorisanje slike.