



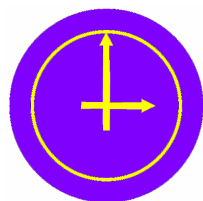
Informatika

Uvodni deo 3. čas

Rudarsko-geološki fakultet
Rudarski odsek

Informacija i podatak

- Podatak je nosilac informacije u određenom formatu
- Informacija je rezultat organizovanja i obrade podataka, odnosno rukovanja podacima na način koji ima za cilj da doprinese znanju primaoca
- Formati podataka koji nose istu informaciju mogu biti različiti



03:00:00

- U računarskim sistemima informacije se predstavljaju podacima u binarnom formatu
- Na računarskim uređajima vidljivim za korisnika takve informacije se uobličavaju kao podaci čitljivi za korisnika



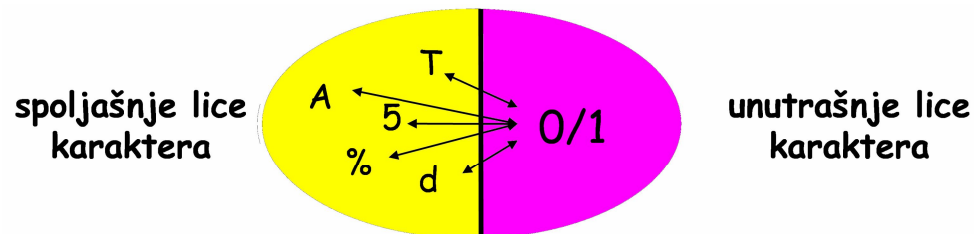
Glif

- U memoriji računara se nalaze isključivo binarne cifre organizovane u bajtove ili reči ali bi unos podataka u računar u binarnom obliku od strane korisnika bio teško zamisliv
- "Komunikacija" između korisnika i računara se odvija upotrebom uobičajenih simbola: slova, cifara, interpunkcijskih i drugih specijalnih znakova
- Ovi znaci imaju određeni grafički oblik koji se naziva **glif** – to je "šara" koju prepoznajemo kao određeni znak u određenom sistemu pisanja



Karakter

- Simbole koji se unose u računar korisnik vidi kao glif dok se oni u računaru predstavljaju kao određena binarna kombinacija
- Ova dvostruka reprezentacija jednog simbola se naziva **karakter**
- Na taj način karakter ima svoje spoljašnje lice (glif), okrenuto ka korisniku, i unutrašnje lice (binarnu reprezentaciju) nad kojom radi računar





Karakterski skup

- Skup karaktera koji se koristi na određenom računarskom sistemu nazivamo **karakterski skup**
- Izbor glifova kojima će biti pridružene binarne kombinacije kao i povezivanje glifa i određene binarne kombinacije je u načelu proizvoljno:
 - ne postoje nikakva unapred zadata ograničenja po kojima određenoj binarnoj kombinaciji mora da odgovara određeni glif



ASCII KOD

- I ako ne postoje unapred zadata ograničenja ipak se za **karacterske skupove utvrđuju određeni kriterijumi i zahtevi** da bi se time omogućila razmena podataka između različitih korisnika
- Kriterijumi i zahtevi za karacterske skupove se mogu definisati na različite načine čime se definišu različite šeme kodiranja ili **kodovi**
- Jedan od prvih je karacterski skup ASCII (*American Standard Code for Information Interchange*) koji se još uvek vrlo široko koristi
- Po ovom standardu karakteri se predstavljaju pomoću 7 bitova odnosno jednog bajta u kome je prvi bit uvek 0



Kontrolni i grafički karakteri

- **Kontrolni karakteri ASCII koda (od 0 do 31):**
 - namenjeni uređivanju podataka
 - na primer, *carriage return* (povratak sa tekuće pozicije u redu na početak reda) ili *line feed* (prelazak na novi red)
- **Grafički karakteri ASCII koda (32 do 127):**
 - namenjeni predstavljanju određenog skupa glifova
 - izbor glifova je izvršen u skladu sa osnovnim potrebama engleskog jezika
 - čine ga velika i mala slova engleskog alfabeta, cifre dekadnog brojnog sistema i izvestan broj specijalnih simbola kao što su = + , . itd.

Grafički karakteri u ASCII kodu

32 (razmak)	48 0	64 @	80 P	96 `	112 p
33 !	49 1	65 A	81 Q	97 a	113 q
34 "	50 2	66 B	82 R	98 b	114 r
35 #	51 3	67 C	83 S	99 c	115 s
36 \$	52 4	68 D	84 T	100 d	116 t
37 %	53 5	69 E	85 U	101 e	117 u
38 &	54 6	70 F	86 V	102 f	118 v
39 '	55 7	71 G	87 W	103 g	119 w
40 (56 8	72 H	88 X	104 h	120 x
41)	57 9	73 I	89 Y	105 i	121 y
42 *	58 :	74 J	90 Z	106 j	122 z
43 +	59 ;	75 K	91 [107 k	123 {
44 ,	60 <	76 L	92 \	108 l	124
45 -	61 =	77 M	93]	109 m	125 }
46 .	62 >	78 N	94 ^	110 n	126 ~
47 /	63 ?	79 O	95 _	111 o	127



Kolaciona sekvencija

- Predstavlja redosled karaktera prema rastućim vrednostima brojeva pridruženih glifovima
- Kolaciona sekvencija definiše jedan poredak među *elementima karakterskog skupa*
- Na primer, u odnosu na kolacionu sekvenciju u ASCII kodu je $A < B < a < b$ jer za brojeve koji su im pridruženi važi: $65 < 66 < 97 < 98$



Konstrukcija ASCII koda

- Raspored karaktera u ASCII kodu u odnosu na kolacionu sekvenciju ima sledeće karakteristike:
 - razmak (ili blanko) prethodi ciframa i slovima
 - velika i mala slova engleskog alfabeta kao i cifre dekadnog sistema su u rastućem redosledu i u neprekinutom (kontingentnom) nizu
 - mala slova pomaknuta su za 32 pozicije u odnosu na velika (npr. veliko slovo **B** je na poziciji 66, a malo slovo **b** na poziciji $66+32=98$).



Nacionalne verzije

- Zapis tekstova na velikom broju evropskih jezika pored slova predviđenih **ASCII**-standardom zahteva i dodatna slova kao što su *ć, č, š, ž, đ* ili *ß* za nemački, *ñ* za španski *å* za švedski i druge nordijske jezike...
- U nacionalnim verzijama ASCII koda (**ISO/IEC 646** standard) za specifičnosti nacionalnih alfabeta određene su pozicije: 64, 91-94, 96, 123-126
- **ISO** - International Organization for Standardization (definiše međunarodne standarde u svim oblastima)
- **IEC** - International Electrotechnical Commission (odgovorna za standardizaciju elektronske opreme)

YU-ASCII I (latinični)

yu - ascii (latinični)					
64	@	Ž	96	`	ž
91	[Š	123	{	š
92	\	Đ	124		đ
93]	Ć	125	}	ć
94	^	Č	126	~	č

- U YU-ASCII I -kodu ne važi uslov da se slova nalaze u rastućem redosledu i kontigentnom nizu u odnosu na kolacionu sekvenciju: Ž < A < ... < Z < Š < Đ < Č < Ć



YU-ASCII (ćirilični)

- Ćirilično pismo zahteva još i pozicije za glifove slova Љ, Њ i Џ
- Oni su u odnosu na ASCII -kod postavljene na pozicije glifova Q, W i X

yu - ascii - ćirilični					
81	Q	Љ	113	q	Љ
87	W	Њ	119	w	Њ
88	X	Џ	120	x	Џ



ISO/IEC 8859 kod

- Standard za 8-bitno kodiranje karaktera
- Sadrži 15 delova: ISO/IEC 8859-1, ISO/IEC 8859-2, itd. od kojih se svaki može smatrati standardom za sebe
- Rešava problem nacionalnih verzija tako što koristi osmi bit u bajtu čime se dobija još 128 karaktera u koje se smeštaju slova nacionalnih alfabeta
- Sadrži desetak varijanti latinice
- Sadrži varijante za ćirilicu, grčki, arapski i hebrejski...
- Prvih 127 karaktera isti kao i u ASCII kodu



ISO/IEC 8859-2

Latinica-2 (Centralno-Evropski)

- Zvanično ISO/IEC 8859-2 ili manje zvanično Latin-2, sadrži 191 latiničnih slova kodiranih osmobiitnim vrednostima
- Koristi se za razmenu podataka na sledećim evropskim jezicima: češki, mađarski, poljski, rumunski, srpski (latinično), slovački, slovenački, bošnjački, hrvatski, gornjesorbski i donjesorbski (jezici lužičkih Srba ili Soraba, slovenske manjine u oblasti Lusatia u nemačkim državama Saksonija i Brandenburg)
- Može se koristiti i za nemački i engleski (koji nominalno koriste ISO 8859-1)

Tabela ISO/IEC 8859-2 karaktera

	x0	x1	x2	x3	x4	x5	x6	x7	x8	x9	xA	xB	xC	xD	xE	xF
0x	<i>neiskorišćeno</i>															
1x																
2x	SP	!	"	#	\$	%	&	'	()	*	+	,	-	.	/
3x	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4x	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5x	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
6x	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7x	p	q	r	s	t	u	v	w	x	y	z	{		}	~	
8x	<i>neiskorišćeno</i>															
9x																
Ax	NBSP	Ą	ˆ	Ł	ą	ł	Ś	ś	˝	Š	š	Ť	ž	-	Ž	Ž
Bx	°	ą	ˆ	ł	ą	ł	ś	ˆ	˝	š	š	ť	ž	˝	ž	ž
Cx	Ŕ	Á	Â	Ă	Ä	Í	Ć	Ç	Č	É	Ę	Ě	Ě	Í	Î	Ď
Dx	Đ	Ń	Ň	Ó	Ô	Õ	Ö	×	Ř	Ů	Ú	Ů	Ü	Ý	Ť	ß
Ex	í	á	â	ă	ä	í	ć	ç	č	é	ę	ě	ě	í	î	ď
Fx	đ	ń	ň	ó	ô	õ	ö	÷	ř	ů	ú	ů	ü	ý	ť	·



ISO/IEC 8859-5 - Latinica/Ćirilica

- Zvanično ISO/IEC 8859-5 ili manje zvanično Ćirilica
- Sadrži 52 latinična slova i 92 ćirilična slova kodirana osmobitnim vrednostima
- Koristi se za razmenu podataka na evropskim jezicima koji koriste ćiriličnu azbuku: bugarski, beloruski, ruski, ukrajinski i srpski

ISO/IEC 8859-5 Latinica/Ćirilica

	x0	x1	x2	x3	x4	x5	x6	x7	x8	x9	xA	xB	xC	xD	xE	xF
0x	<i>neiskorišćeno</i>															
1x																
2x	SP	!	"	#	\$	%	&	'	()	*	+	,	-	.	/
3x	0	1	2	3	4	5	6	7	8	9	:	;	<	=	r	?
4x	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5x	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
6x	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7x	p	q	r	s	t	u	v	w	x	y	z	{		}	~	
8x	<i>neiskorišćeno</i>															
9x																
Ax	NBSP	Ě	Ђ	Ѓ	Є	Ѕ	І	Ї	Ј	Љ	Њ	Ћ	Ќ	-	Ў	Џ
Bx	А	Б	В	Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О	П
Cx	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Ь	Э	Ю	Я
Dx	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п
Ex	р	с	т	у	ф	х	ц	ч	ш	щ	ъ	ы	ь	э	ю	я
Fx	№	ě	ђ	ѓ	є	ѕ	і	ї	ј	љ	њ	ћ	ќ	џ	ў	џ



Unicode

- Industrijski standard čiji je cilj da se obezbedi mogućnost kodiranja tekstova u svim formatima i na svim jezicima
- Projekat čija je ambicija da zameni sve postojeće šeme kodiranja koje imaju svoja ograničenja kada su u pitanju višejezična okruženja
- Najveća i najkompletnija šema kodiranja karaktera koja je postala dominantna metoda u internacionalizaciji i lokalizaciji softvera (adaptaciji softvera za okruženja iz kojih nisu potekli - za jezike drugih nacija i kultura)
- Standard se implementira i u mnogim novim tehnologijama kao što su XML, programski jezik Java i savremeni operativni sistemi
- Sadrži više 107.000 karaktera za 90 različitih vrsta pisama



Prikazivanje karaktera u Unicode-u

- Karakteri se zapisuju pomoću jednog ili više bajtova
- Ima nekoliko varijanti od kojih se najčešće koriste UTF-8 (*Unicode Transformation Format*) i UTF-16
- Prvih 256 karaktera je isto kao u ISO 8859-1 kodu (a time i prvih 127 isto kao u ASCII kodu) kako bi se tekstovi pisani u ovim kodovima mogli neposredno konvertovati
- Mnogo u suštini istih karaktera kodirano je više puta različitim kodovima
- Sadrži unapred kombinovane verzije svih uobičajenih varijanti slovo/dijakirik ("akcent") ali dozvoljava i kombinovanje karaktera: na primer ć se može predstaviti i kao malo latinično slovo c iza koga stoji uzlazni akcent
- Dozvoljava prikazivanje karaktera u oba smera da bi se omogućilo da se tekstovi u jezicima kao što su arapski i hebrejski prikazuju sdesna na levo