

UNDEFINED CHARACTER

Diana Šantavec
diana.santavec@gmail.com

Istraživačka stanica Petnica

07.04.2024.

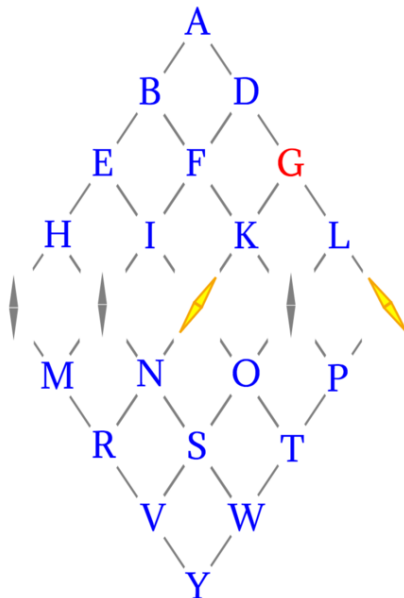


- Šta je ovo predavanje?

Needle Telegraph



Needle Telegraph



International Morse Code

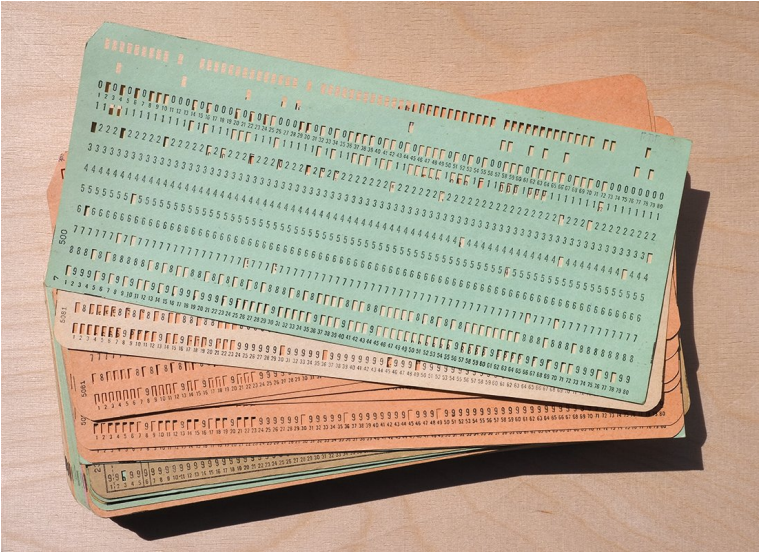
A	· -	N	- ·	1	· - - - -
B	- · · ·	O	- - -	2	· - - - -
C	- · - ·	P	· - - ·	3	· · - - -
D	- · ·	Q	- - · -	4	· · · -
E	·	R	· - ·	5	· · · ·
F	· · - ·	S	· · ·	6	- · · · ·
G	- - ·	T	-	7	- - · · ·
H	· · · ·	U	· · -	8	- - - · ·
I	· ·	V	· · · -	9	- - - - ·
J	· - - -	W	· - -	0	- - - - -
K	- · -	X	- · · -	.	· - · · · ·
L	· - · ·	Y	- · - -	,	- - · · - -
M	- -	Z	- - · ·	?	· · · · ·

- **Ali pre toga**
- Teletype computers
- Bušene kartice

Teletype Computers



Bušene kartice



ASCII

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

www.alpharhms.com

```
#include <stdio.h>

int main(){
    char test;
    test = 'a';

    printf("%c", test);
    printf("\n");
    printf("%d", test);
}
```

Ispis karaktera

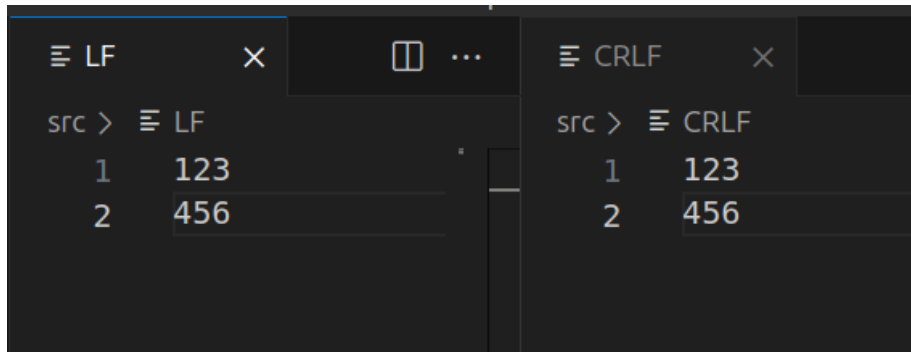
```
dianas@dianas-ThinkPad ~/P/P/P/code (master)>  
dianas@dianas-ThinkPad ~/P/P/P/code (master)> gcc print\ text.c -o print\ text  
dianas@dianas-ThinkPad ~/P/P/P/code (master)> ./print\ text  
a  
97  
dianas@dianas-ThinkPad ~/P/P/P/code (master)> █
```

Kontrolni karakteri

0	00000000	NUL	Null char
1	00000001	SOH	Start of Heading
2	00000010	STX	Start of Text
3	00000011	ETX	End of Text
4	00000100	EOT	End of Transmission
5	00000101	ENQ	Enquiry
6	00000110	ACK	Acknowledgment
7	00000111	BEL	Bell
8	00001000	BS	Back Space
9	00001001	HT	Horizontal Tab
10	00001010	LF	Line Feed
11	00001011	VT	Vertical Tab
12	00001100	FF	Form Feed
13	00001101	CR	Carriage Return
14	00001110	SO	Shift Out / X-On
15	00001111	SI	Shift In / X-Off

16	00010000	DLE	Data Line Escape
17	00010001	DC1	Device Control 1
18	00010010	DC2	Device Control 2
19	00010011	DC3	Device Control 3
20	00010100	DC4	Device Control 4
21	00010101	NAK	Negative Ack
22	00010110	SYN	Synchronous Idle
23	00010111	ETB	End of Transmit Block
24	00011000	CAN	Cancel
25	00011001	EM	End of Medium
26	00011010	SUB	Substitute
27	00011011	ESC	Escape
28	00011100	FS	File Separator
29	00011101	GS	Group Separator
30	00011110	RS	Record Separator
31	00011111	US	Unit Separator

LF vs CRLF



LF vs CRLF

```
dianas@dianas-ThinkPad ~/P/P/b/src (master)>
dianas@dianas-ThinkPad ~/P/P/b/src (master)> hexdump -C LF
00000000  31 32 33 0a 34 35 36                                |123.456|
00000007
dianas@dianas-ThinkPad ~/P/P/b/src (master)> hexdump -C CRLF
00000000  31 32 33 0d 0a 34 35 36                                |123..456|
00000008
dianas@dianas-ThinkPad ~/P/P/b/src (master)>
dianas@dianas-ThinkPad ~/P/P/b/src (master)\
```

ASCII - brojevi

48	30	060	0110000	0
49	31	061	0110001	1
50	32	062	0110010	2
51	33	063	0110011	3
52	34	064	0110100	4
53	35	065	0110101	5
54	36	066	0110110	6
55	37	067	0110111	7
56	38	070	0111000	8
57	39	071	0111001	9

ASCII - slova

65	41	101	1000001	A
66	42	102	1000010	B
67	43	103	1000011	C
68	44	104	1000100	D

■ ■ ■

89	59	131	1011001	Y
90	5A	132	1011010	Z

97	61	141	1100001	a
98	62	142	1100010	b
99	63	143	1100011	c
100	64	144	1100100	d

■ ■ ■

121	79	171	1111001	y
122	7A	172	1111010	z

Šta je sa osmim bitom?



I super, nemamo problema :D

I super, nemamo problema :D

Zar ne?

Problemi na različitim sistemima

- Unicode Transformation Format - 8 bitova
- Prvih 128 karaktera je isto kao i za ASCII
- replacement character

- combining character

a + ö + ã + ¨ + ^ → ã

0061 0308 0303 0323 032D

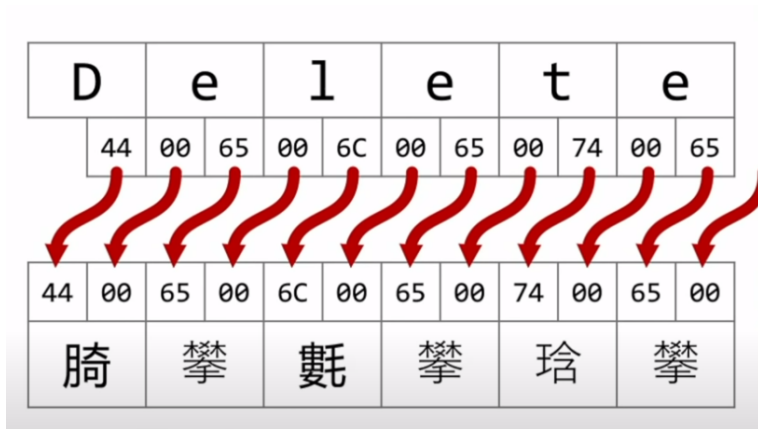
UTF-8 encoding

Code point <-> UTF-8 conversion

First code point	Last code point	Byte 1	Byte 2	Byte 3	Byte 4
U+0000	U+007F	0xxxxxxx			
U+0080	U+07FF	110xxxxx	10xxxxxx		
U+0800	U+FFFF	1110xxxx	10xxxxxx	10xxxxxx	
U+10000	[nb 2] U+10FFFF	11110xxx	10xxxxxx	10xxxxxx	10xxxxxx

Hakovani smo!?

- UTF16



Kako imamo smajlije?

☒=WOBTAX

☒=OBTAX

☒=VICTORY

☒=VICTOR

☒=VICTORina

☒=+ov' "

☒=O*\$

☒=Yov

☒=TYUOV

☒=O*\$o

- Japan
- Oxford uvodi u rečnik - 2013

Emoji

```
var 😊 = "Smiley"
println(😊) // will print "Smiley"
let 🌐 = "🐶🐺🐱🐭"
var 🚢: String[] = []
for ❤️ in 🌐 {
    🚢.append(❤️+❤️)
}
println(🚢) // will print [🐶🐶, 🐺🐺, 🐱🐱, 🐭🐭]
```



Da li imamo probleme sa platformama?

Apple



LG



Google



Samsung



Microsoft



HTC



Ipak da...



Spajanje emoji-a



U+1F469 U+1F3FE

F0 9F 91 A9 F0 9F 8F BF

Spajanje emoji-a nije uvek očigledno



		
U+1F469 Woman	U+200D Zero Width Joiner	U+1F680D Rocket

Kako pravimo piratsku zastavu?



Kako pravimo piratsku zastavu?



U+2620 + U+FE0F


Kako pravimo piratsku zastavu?



🚩 U+1F3F4 U+200D 🦴 U+2620 U+FE0F

Regional Indicator Symbol

- Par simbola identifikuje emoji zastavu

Regional indicator symbols subset of Enclosed Alphanumeric Supplement ^[1] Official Unicode Consortium code chart  (PDF)																
	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
...	(U+1F100-U+1F1E5 omitted)															
U+1F1Ex							A	B	C	D	E	F	G	H	I	J
U+1F1Fx	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Notes 1. ^ As of Unicode version 14.0																

Kraj?

Kraj?

- Plain Text - Dylan Beattie - NDC Oslo 2021
- Emoji how do they even work - MacLemon

Pitanja?