

# Pracownia komputerowa

Dariusz Wardecki, wyk. IX

# Powtórzenie

Ile bajtów zawiera tekst (ASCII)?

```
Pracownia komputerowa  
jest najciekawsza!
```

# Kod ASCII

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

# Standardy ISO-8859

## ISO-8859-2

- Znaki wykorzystywane w językach środkowoeuropejskich.
- Kody 128 . . . 255 przypisane innym znakom, niż w ISO-8859-1.

## ISO-8859-15

Rewizja ISO-8859-1 wprowadzająca znak waluty euro.

Zasada „1 znak – 1 bajt” powoduje problemy z przenoszeniem tekstów między systemami wykorzystywanymi w różnych krajach.

# Standard Unicode

## ISO/IEC 10646:2003, Universal Character Set (UCS)

- 1 Koniec zasady „1 znak – 1 bajt” .
- 2 Znaki reprezentowane przez kody wielobajtowe.
- 3 „Rozszerzenie” ASCII (znaki ASCII odpowiadają kodom 0 . . . 127).
- 4 Brak zgodności z ISO-8859 i stronami kodowymi IBM oraz Microsoft.

## UTF-8 (8-bit Unicode Transformation Format)

- Jednobajtowe reprezentacje znaków ASCII.
- Dwubajtowe reprezentacje znaków z języków europejskich.
- Najbardziej popularny format Unicode.

Dzięki Unicode możliwe jest tworzenie **uniwersalnych** dokumentów tekstowych (tzn. tekstów wyświetlanych wszędzie tak samo).

# Kodowanie UTF-8

Bity	Największa wartość	Bajt 1	Bajt 2	Bajt 3	Bajt 4
7	U+007F	0xxxxxxx			
11	U+07FF	110xxxxx	10xxxxxx		
16	U+FFFF	1110xxxx	10xxxxxx	10xxxxxx	
21	U+1FFFFFF	11110xxx	10xxxxxx	10xxxxxx	10xxxxxx

# Kodowanie UTF-8

Przykład:

Znak		Kod	Bin	Hex
@	U+0040	100 0000	0100 0000	40
À	U+0104	001 0000 0100	1100 0100 1000 0100	c4 84
龜	U+2EF1	0010 1110 1111 0001	1110 0010 1011 1011 1011 0001	e2 bb b1
𠄎	U+10304	0 0001 0000 0011 0000 0100	1111 0000 1001 0000 1000 1100 1000 0100	f0 90 8c 84

# Pliki tekstowe

## Plik tekstowy (*ang. text file*)

Zawiera dane zakodowane w postaci kodów znaków, które powinny być interpretowane zgodnie z ustalonym zbiorem znaków.

- 1 Tzw. **znaki sterujące** (*ang. control characters*) odpowiadają specjalnym kodom określającym sposób **formatowania wydruku** (np. znaki ASCII o kodach 0 – 31).
- 2 Pliki tekstowe nie zawierają informacji o tym jaki zbiór znaków jest odpowiedni do interpretacji ich zawartości (program drukujący tekst lub użytkownik powinien to określić).

## Pliki tekstowe mają ograniczone zastosowania

- 1 Mogą zawierać tylko tekst (brak grafiki, tabel, odnośników itp.).
- 2 Nie zawierają informacji o kroju i wielkości czcionki, kolorach itp.



# Pliki metatekstowe

## Plik metatekstowe

Zawiera dane w postaci tekstu (tzn. kodów znaków jak dla „zwykłego” pliku tekstowego), z których część może być interpretowana jako **znaczniki** (*ang. tag*) określające m. in.

- Oczekiwany wygląd dokumentu (np. krój i wielkość czcionki itp.).
- Jego powiązania z innymi dokumentami (np. odnośniki).
- Zbiór znaków, zgodnie z którym mają być interpretowane kody znaków w dokumencie.

## Reguły tworzenia i interpretacji znaczników określają metajęzyki

- HTML (*ang. HyperText Markup Language*)
- XML (*ang. Extensible Markup Language*)
- SGML (*ang. Standard Generalized Markup Language*)

# Postscript

- Uniwersalny język opisu strony opracowany przez Adobe
- Pliki postscriptowe najczęściej mają rozszerzenie .ps lub .eps
- Dokumenty ps można tworzyć m.in. w programach Latex lub Gnuplot

## Przykład

```
%!  
/Helvetica findfont 72 scalefont setfont  
72 72 moveto  
(Hello, world!) show  
showpage
```

# Grafika rastrowa i wektorowa

## Grafika rastrowa (*ang. raster graphics*)

Obraz jest reprezentowany przez (ogólnie) prostokątną siatkę punktów o różnych kolorach, czyli tzw. **pikseli** (*ang. pixel, picture element*). Oczka tej siatki **nie muszą** mieć jednakowej szerokości i wysokości.

## Mapa bitowa, bitmapa (*ang. bitmap*)

Struktura danych reprezentująca obraz w postaci gotowej do wydrukowania (lub wyświetlenia). Kolor każdego piksela reprezentuje określona liczba bitów danych.

## Grafika wektorowa (*ang. vector graphics*)

Obraz jest reprezentowany z pomocą **instrukcji**, które mają być wykonane w celu utworzenia go (np. „narysuj linię prostą”, „narysuj koło” itp.). Poszczególne punkty identyfikuje się podając ich **współrzędne** (wektory).

# Kolory w grafice rastrowej

## RGB (*ang. red, green, blue*)

Kolor piksela określa się podając natężenie każdej z trzech podstawowych barw światła (czerwony, zielony, niebieski). Zwykle jeden piksel jest reprezentowany przez 3 lub 4 bajty danych. Reprezentacja odpowiednia dla monitorów i wyświetlaczy LCD, OLED itp.

## CMYK (*ang. cyan, magenta, yellow, black*)

Kolor piksela określa się podając natężenie każdego z czterech kolorów „farb” (turkusowy, purpura, żółty, czarny). Reprezentacja bardziej odpowiednia do drukowania na papierze.

# Formaty plików grafiki rastrowej

- BMP (*ang. bitmap*)
  - Istnieją co najmniej 2 formaty plików graficznych o takiej nazwie.
- PBM (*ang. Portable Bitmap*)
  - Używany w systemach wykorzystujących środowisko X Windows.
- TIFF (*ang. Tagged Image File Format*), ISO 12639
- GIF (*ang. Graphics Interchange Format*), kompresja LZW
  - 8 bitów na piksel, 24-bitowa przestrzeń barw (palety referencyjne).
  - Od 1994 w „szarej strefie” z powodu patentu na algorytm kompresji (Unisys).
- PNG (*ang. Portable Network Graphics*), ISO/IEC 15948:2004, kompresja bezstratna
- JPEG (*ang. Joint Photographic Experts Group*), ISO/IEC 10918, kompresja stratna

# Grafika wektorowa

Nie tylko grafika, także dokumenty tekstowe itp. (praktycznie wszystkie rodzaje dokumentów).

- PostScript (PS) – język programowania wykorzystywany jako język opisu stron (*ang. page description language*) dla ploterów, drukarek itp. (1982).
- EPS (*ang. Encapsulated PostScript*) – kod w języku PostScript z dołączonym „podglądem” w niskiej rozdzielczości.
- PDF (*ang. Portable Document Format*), ISO/IEC 32000-1:2008 – opis strony z „dodatkami”, jak odnośniki (*ang. link*) itp., („przetworzony” PostScript).
- SVG (*ang. Scalable Vector Graphics*)
- DjVu (bardzo dobra kompresja zeskanowanych dokumentów).

# Dokumenty z możliwością wprowadzania zmian

Zawartość pliku reprezentuje dokument

- Sformatowany tekst.
- Ewentualnie grafika, tabele itp.
- Zawiera informacje o sposobie wyświetlania lub drukowania.
- Może być skompresowany.

W przeszłości istniało wiele konkurencyjnych zamkniętych formatów

- 1 Brak przenośności dokumentów dokuczliwy dla użytkowników.
- 2 Niezgodności między różnymi wersjami tego samego programu (!).
- 3 Nieformalnym standardem stały się formaty z MS Office 97:
  - DOC – dokumenty tekstowe.
  - XLS – arkusze kalkulacyjne.
  - PPT – prezentacje.

# Open Document Format

ISO/IEC 26300:2006

Open Document Format for Office Applications (OpenDocument) v1.0.

Zestaw otwartych formatów plików, opartych na XML (z kompresją), dla różnych rodzajów dokumentów:

- ODT – dokumenty tekstowe.
- ODS – arkusze kalkulacyjne.
- ODP – prezentacje.
- ODG – rysunki (grafika wektorowa).
- ODF – wzory matematyczne.

Większość dostępnych na rynku pakietów biurowych wspiera ODF.

- Dostępna jest „wtyczka” dla Microsoft Office.