

Dane i ich kodowanie

R. Robert Gajewski
Wydział Inżynierii Lądowej



Liczby całkowite

Liczby naturalne

- Kolejna kombinacja
- Z definicji – „potęgi”

Liczba naturalna	0	1	2	3	4	5	6	7
Wartość binarna	000	001	010	011	100	101	110	111

Pozycja/potęga	7	6	5	4	3	2	1	0
Dwa do...	128	64	32	16	8	4	2	1
Bit	1	0	1	0	1	0	1	0
$128+32+8+2=$	128	0	32	0	8	0	2	0

Kod z bitem znaku

- Najstarszy, skrajny lewy bit to bit znaku
- 1 to liczba ujemna, 0 dodatnia
- Zero można zakodować na dwa sposoby
 - Dodatni 000
 - Ujemny 100

Liczba naturalna	0	1	2	3	4	5	6	7
Wartość binarna	000	001	010	011	100	101	110	111
Interpretacja z bitem z.	0	1	2	3	-0	-1	-2	-3

Kod z uzupełnieniem do 1

- Najstarszy bit ma wartość zero
- Liczba ujemna to bitowa negacja liczby dodatniej
- Niejednoznaczność kodowania zera
 - Dodatnie 000
 - Zanegowane 111

Liczba naturalna	0	1	2	3	4	5	6	7
Wartość binarna	000	001	010	011	100	101	110	111
Interpretacja z bitem z.	0	1	2	3	-3	-2	-1	-0

Kod z uzupełnieniem do 2

- Liczbę przeciwną dla n bitów otrzymujemy odejmując daną liczbę od 2^n
- Zamiana liczby na U2
 - Inwersja bitów – 0 na 1 a 1 na 0
 - Zwiększamy wynik o 1

Liczba naturalna	0	1	2	3	4	5	6	7
Wartość binarna	000	001	010	011	100	101	110	111
Interpretacja z bias	0	1	2	3	-4	-3	-2	-1
Inwersja		110	101	100	011			

Kodowanie z przesunięciem

- Każdej liczbie, najczęściej całkowitej, przyporządkowujemy liczbę nieujemną, przez dodanie do zamienianej liczby wartości n .
- Liczbę $-n$ kodujemy jako 0, kolejną liczbę jako 1 i tak dalej.

Liczba naturalna	0	1	2	3	4	5	6	7
Wartość binarna	000	001	010	011	100	101	110	111
Interpretacja z bias	-4	-3	-2	-1	0	1	2	3

Liczby rzeczywiste

Tekst

ASCII [Wiki]

- American Standard Code for Information Interchange) 1963 rok –
 - 7-bitowy kod przyporządkowujący liczby z zakresu 0–127: literom alfabetu angielskiego, cyfrom, znakom przestankowym i innym symbolom oraz poleceniom sterującym.
- Znaki ASCII dzielą się na:
 - drukowane: 95 znaków o kodach 32–126
 - sterujące: 33 znaki o kodach 0–31 i 127

1972 Printer Manual



USASCII code chart

Bits					0 0 0	0 0 1	0 1 0	0 1 1	1 0 0	1 0 1	1 1 0	1 1 1
b ₄	b ₃	b ₂	b ₁	Column	0	1	2	3	4	5	6	7
↑	↑	↑	↑	Row ↓	0	1	2	3	4	5	6	7
0	0	0	0	0	NUL	DLE	SP	0	@	P	`	p
0	0	0	1	1	SOH	DC1	!	1	A	Q	a	q
0	0	1	0	2	STX	DC2	"	2	B	R	b	r
0	0	1	1	3	ETX	DC3	#	3	C	S	c	s
0	1	0	0	4	EOT	DC4	\$	4	D	T	d	t
0	1	0	1	5	ENQ	NAK	%	5	E	U	e	u
0	1	1	0	6	ACK	SYN	&	6	F	V	f	v
0	1	1	1	7	BEL	ETB	'	7	G	W	g	w
1	0	0	0	8	BS	CAN	(8	H	X	h	x
1	0	0	1	9	HT	EM)	9	I	Y	i	y
1	0	1	0	10	LF	SUB	*	:	J	Z	j	z
1	0	1	1	11	VT	ESC	+	;	K	[k	{
1	1	0	0	12	FF	FS	,	<	L	\	l	
1	1	0	1	13	CR	GS	-	=	M]	m	}
1	1	1	0	14	SO	RS	.	>	N	^	n	~
1	1	1	1	15	SI	US	/	?	O	_	o	DEL

https://upload.wikimedia.org/wikipedia/commons/c/cf/USASCII_code_chart.png

Regular ASCII chart

000	(nul)	016	▶ (dle)	032	sp	048	ò	064	@	080	P	096	`	112	p
001	⊙ (soh)	017	◀ (dcl)	033	!	049	1	065	A	081	Q	097	a	113	q
002	⊕ (stx)	018	↕ (dc2)	034	"	050	2	066	B	082	R	098	b	114	r
003	♥ (etx)	019	!! (dc3)	035	#	051	3	067	C	083	S	099	c	115	s
004	♦ (eot)	020	⌘ (dc4)	036	\$	052	4	068	D	084	T	100	d	116	t
005	♣ (enq)	021	§ (nak)	037	%	053	5	069	E	085	U	101	e	117	u
006	♠ (ack)	022	— (syn)	038	&	054	6	070	F	086	V	102	f	118	v
007	• (bel)	023	⤵ (etb)	039	'	055	7	071	G	087	W	103	g	119	w
008	▣ (bs)	024	↑ (can)	040	(056	8	072	H	088	X	104	h	120	x
009	(tab)	025	↓ (em)	041)	057	9	073	I	089	Y	105	i	121	y
010	(lf)	026	(eof)	042	*	058	:	074	J	090	Z	106	j	122	z
011	♂ (vt)	027	← (esc)	043	+	059	;	075	K	091	[107	k	123	{
012	♀ (np)	028	L (fs)	044	,	060	<	076	L	092	\	108	l	124	
013	(cr)	029	↔ (gs)	045	-	061	=	077	M	093]	109	m	125	}
014	♫ (so)	030	▲ (rs)	046	.	062	>	078	N	094	^	110	n	126	~
015	⊗ (si)	031	▼ (us)	047	/	063	?	079	O	095	_	111	o	127	⊠

Extended ASCII chart

128	Ç	143	Å	158	℞	172	¼	186	∥	200	ℒ	214	∏	228	Σ	242	≥
129	ü	144	É	159	f	173	ı	187	∩	201	∩	215	∏	229	σ	243	≤
130	é	145	æ	160	á	174	«	188	∩	202	∩	216	∏	230	μ	244	∫
131	â	146	Æ	161	í	175	»	189	∩	203	∩	217	∏	231	τ	245	∫
132	ä	147	ô	162	ó	176	⋮	190	∩	204	∩	218	∏	232	Φ	246	÷
133	à	148	ö	163	ú	177	■	191	∩	205	=	219	■	233	⊙	247	≈
134	å	149	ò	164	ñ	178	■	192	∩	206	∩	220	■	234	Ω	248	°
135	ç	150	û	165	Ñ	179	∩	193	∩	207	∩	221	■	235	δ	249	•
136	ê	151	ù	166	ª	180	∩	194	∩	208	∩	222	■	236	∞	250	•
137	ë	152	ÿ	167	º	181	∩	195	∩	209	∩	223	■	237	φ	251	√
138	è	153	Ö	168	¿	182	∩	196	∩	210	∩	224	α	238	ε	252	n
139	ï	154	Ü	169	¬	183	∩	197	∩	211	∩	225	β	239	∩	253	z
140	î	155	ç	170	¬	184	∩	198	∩	212	∩	226	Γ	240	≡	254	■
141	ì	156	£	171	½	185	∩	199	∩	213	F	227	π	241	±	255	
142	Ä	157	¥														

Unicod i UTF [Wiki]

- Komputerowy zestaw znaków mający w zamierzeniu obejmować wszystkie pisma używane na świecie.
 - przydział przestrzeni numeracyjnej poszczególnym grupom znaków
 - sposoby bajtowego kodowania znaków
- UTF - Unicode Transformation Format
- UTF-8 - system kodowania Unicode, wykorzystujący od 1 do 6 bajtów do zakodowania pojedynczego znaku, w pełni kompatybilny z ASCII

UTF-8



Bajty	Bity	B 1	B 2	B 3	B 4
1	7	0xxxxxxx			
2	11	110xxxxx	10xxxxxx		
3	16	1110xxxx	10xxxxxx	10xxxxxx	
4	21	11110xxx	10xxxxxx	10xxxxxx	10xxxxxx
5	26	111110xx
6	31	1111110x

Obrazy

Synteza addytywna [Wiki]

- Zjawisko mieszania barw poprzez sumowanie wiązek światła widzialnego różnych długości.
- **Synteza addytywna** trzech barw przeciwstawnych z koła barw daje światło o barwie białej
 - (teoretycznie, pod warunkiem precyzyjnej przeciwstawności barw i równego natężenia obu strumieni).
- Zachodzi np. podczas projekcji na biały ekran światła ze źródeł o różnych barwach

RGB - [Wiki]

 **R** – red (czerwona)

 **G** – green (zielona)

 **B** – blue (niebieska)

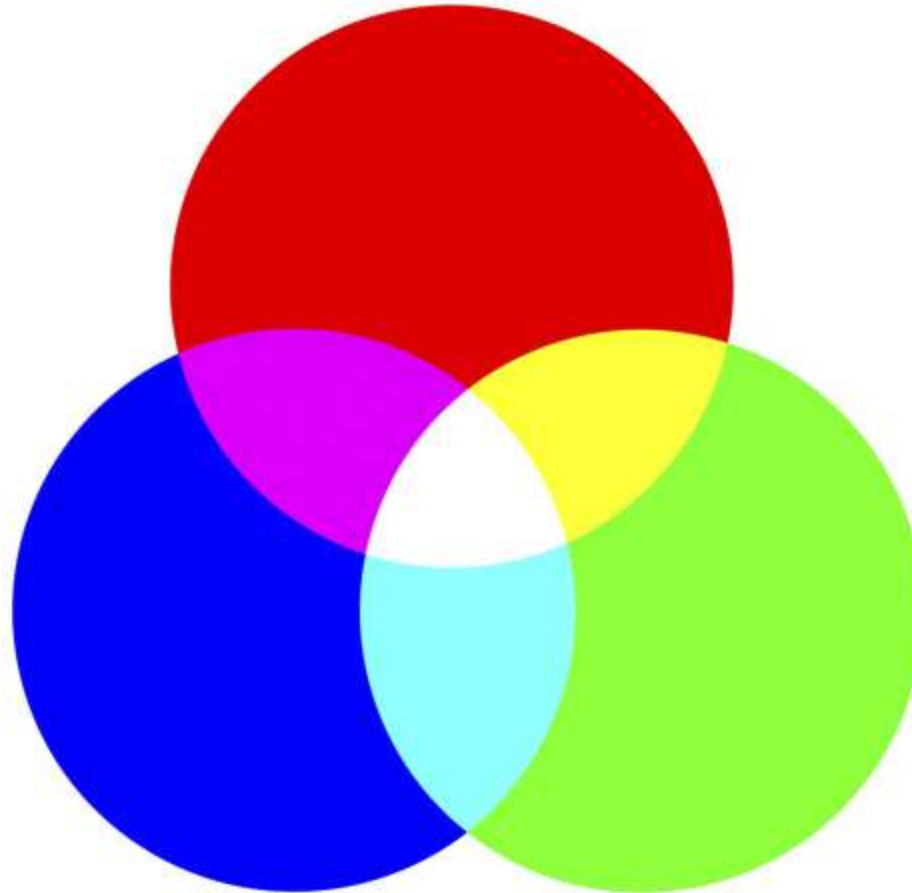
- Model wynikający z właściwości odbiorczych ludzkiego oka, w którym wrażenie widzenia dowolnej barwy można wywołać przez zmieszanie w ustalonych proporcjach trzech wiązek światła o barwie czerwonej, zielonej i niebieskiej

Addytywne mieszanie barw



https://upload.wikimedia.org/wikipedia/commons/2/28/RGB_illumination.jpg

RGB

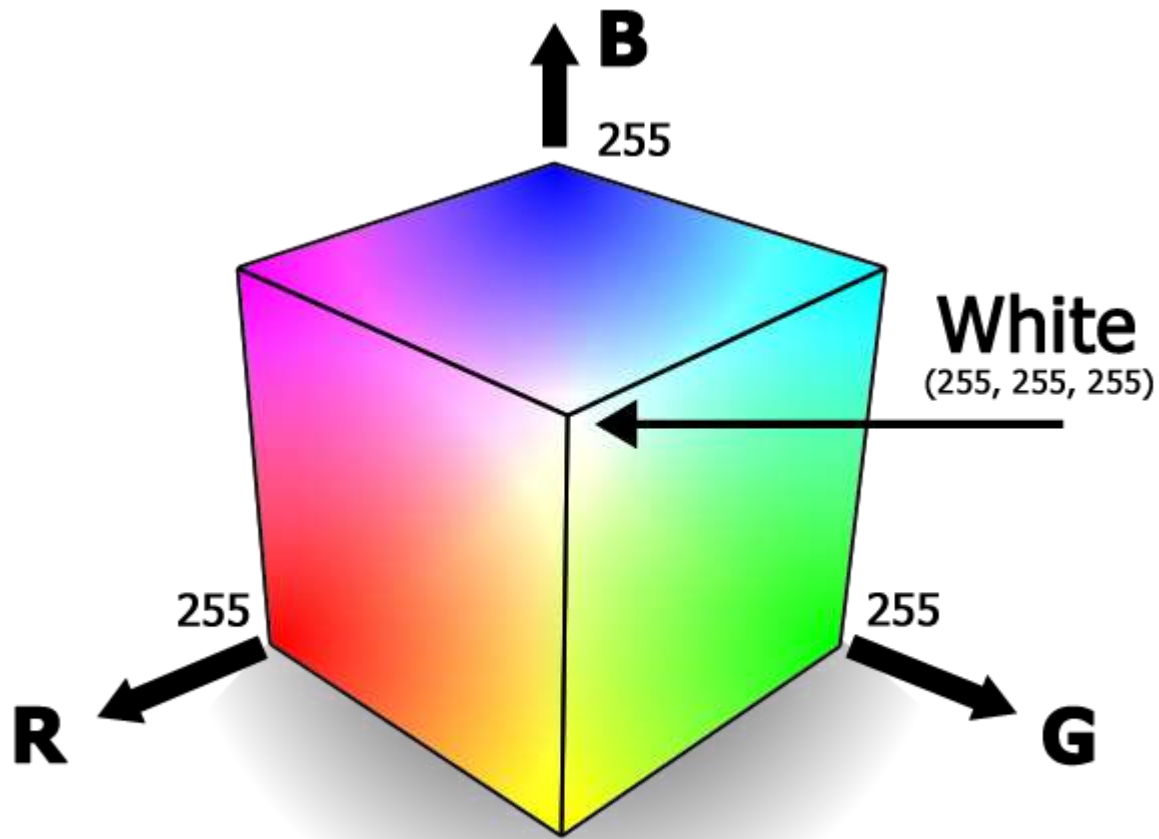


<http://www.tasmyrgb.pl/wp-content/uploads/2014/09/RGB-color-additive-model.jpg>

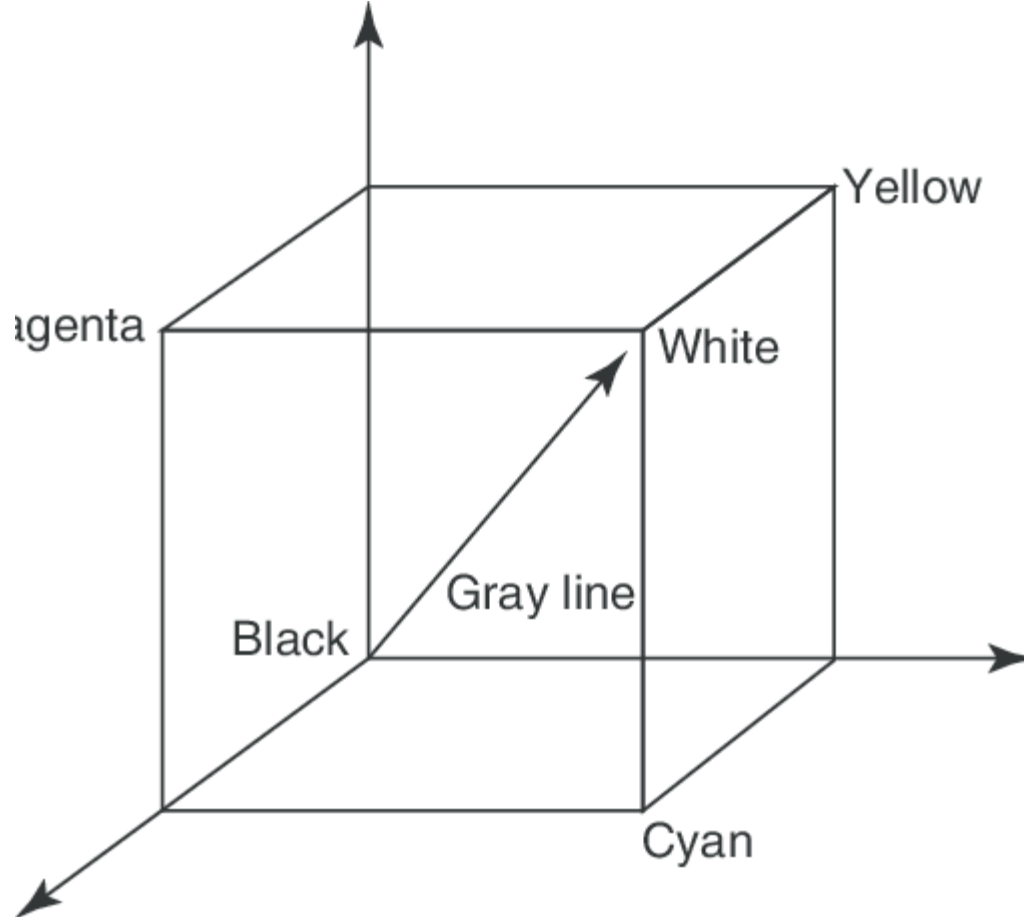
Zapis koloru RGB [Wiki]

- Zapis koloru jako RGB często stosuje się w informatyce (np. palety barw w plikach graficznych, w plikach html).
- Najczęściej stosowany jest 24-bitowy zapis kolorów (po 8 bitów na każdą z barw składowych), w którym każda z barw jest zapisana przy pomocy składowych, które przyjmują wartość z zakresu 0-255.
- W modelu RGB wartość 0 wszystkich składowych daje kolor czarny, natomiast 255 – kolor biały.

Kostka RGB

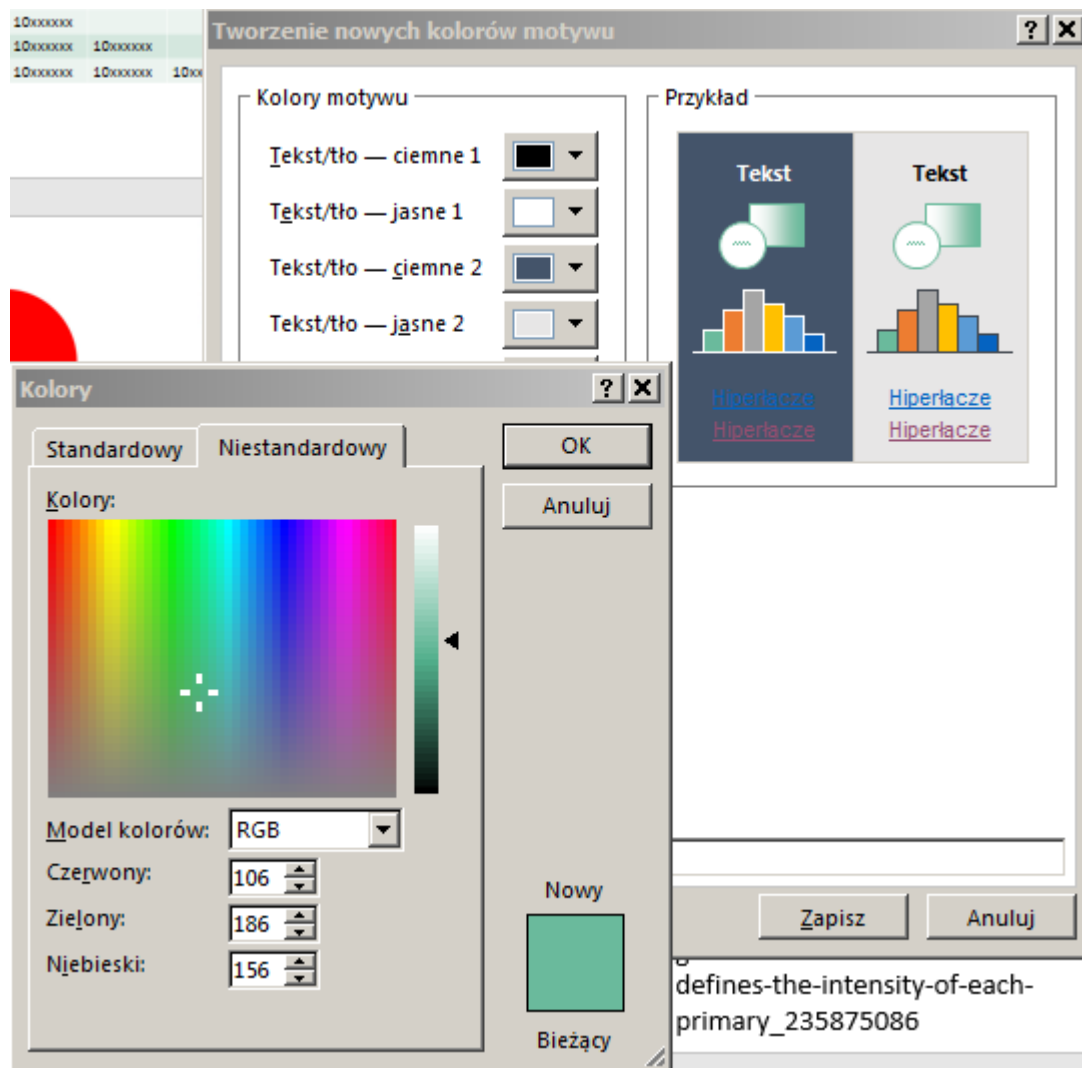


Linia szarości



https://www.researchgate.net/figure/A-red-green-and-blue-RGB-color-cube-The-cube-defines-the-intensity-of-each-primary_235875086

Po co to?



CMYK [Wiki]

- CMYK – zestaw czterech podstawowych kolorów farb drukarskich stosowanych powszechnie w druku wielobarwnym w poligrafii i metodach pokrewnych (atramenty, tonery i inne materiały barwiące w drukarkach komputerowych, kserokopiarkach itp.).
- Na zestaw tych kolorów mówi się również **barwy procesowe** lub **kolory triadowe**

Barwy CMYK



C cyjan (ang. Cyan)



M magenta (ang. Magenta)



Y żółty (ang. Yellow)



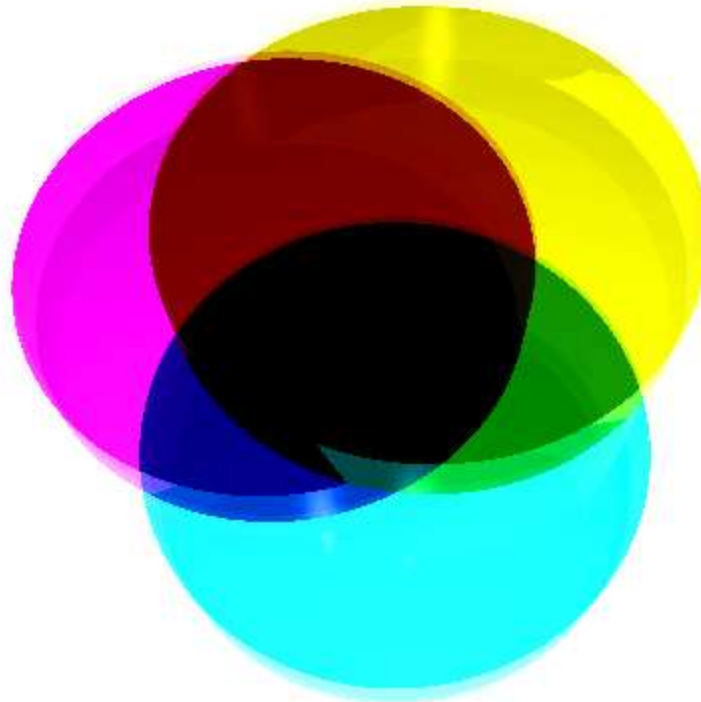
K czarny (ang. Key Colour)

- B zajęte przez Blue w RGB...

Synteza subtraktywna [Wiki]

- Zjawisko mieszania kolorów poprzez odejmowanie promieniowań widzialnych różnych długości (najczęściej poprzez pochłanianie niektórych długości fal przez powierzchnię, od której odbija się światło białe, lub szeregowo ustawionych filtrów światła przechodzącego przez nie).
- **Synteza subtraktywna** zachodzi np. przy mieszaniu farb o różnych kolorach

CMYK



W/ci

Fotografia



https://upload.wikimedia.org/wikipedia/commons/thumb/d/d0/Barns_grand_tetons.jpg/1024px-Barns_grand_tetons.jpg



**Wydział
Inżynierii Lądowej**

Separacja CMY



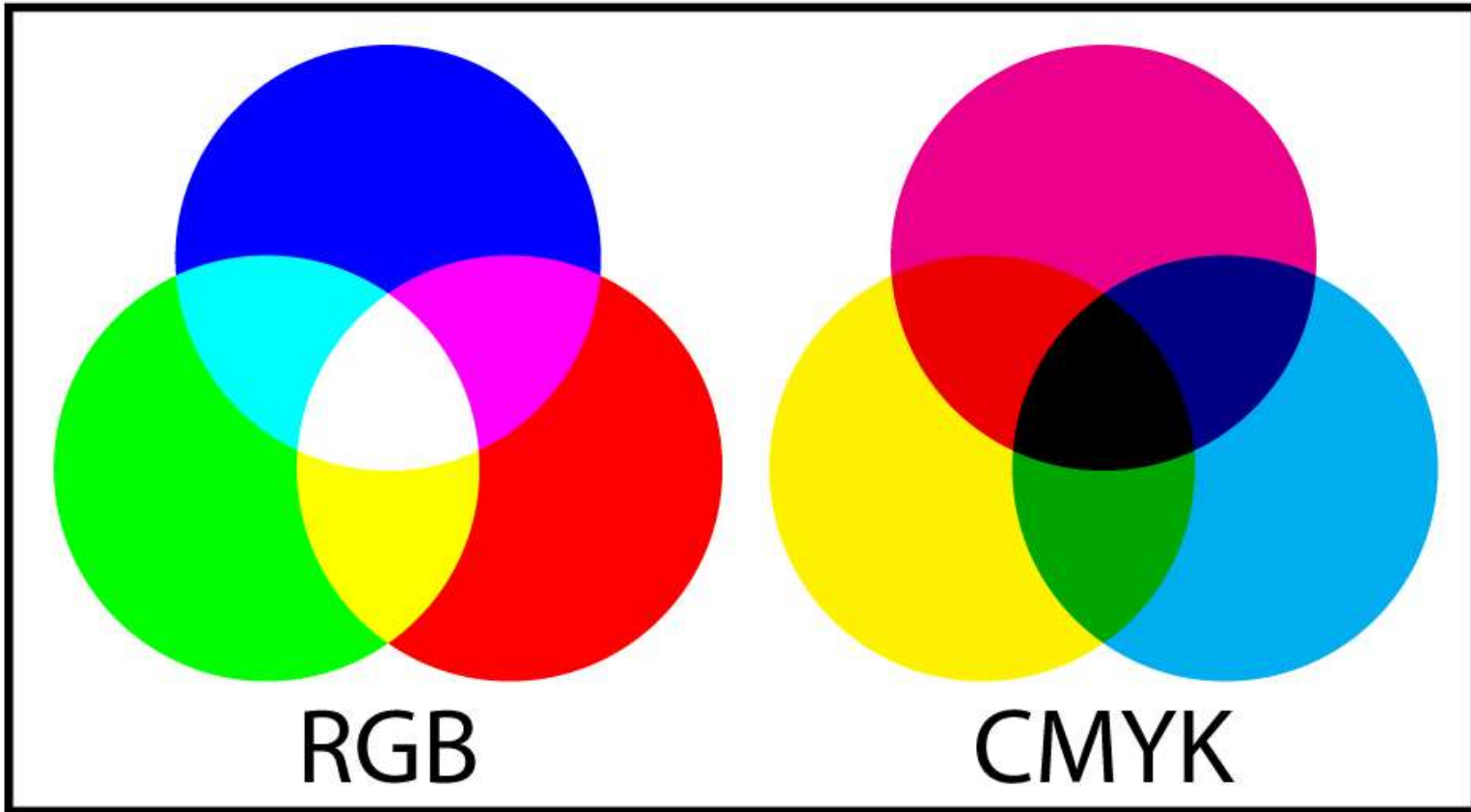
https://upload.wikimedia.org/wikipedia/commons/thumb/f/f0/CMYK_separation_%E2%80%93_93_no_black.jpg/800px-CMYK_separation_%E2%80%93_93_no_black.jpg

Separacja CMYK

https://upload.wikimedia.org/wikipedia/commons/2/2e/CMYK_separation_%E2%80%93_maximum_black.jpg



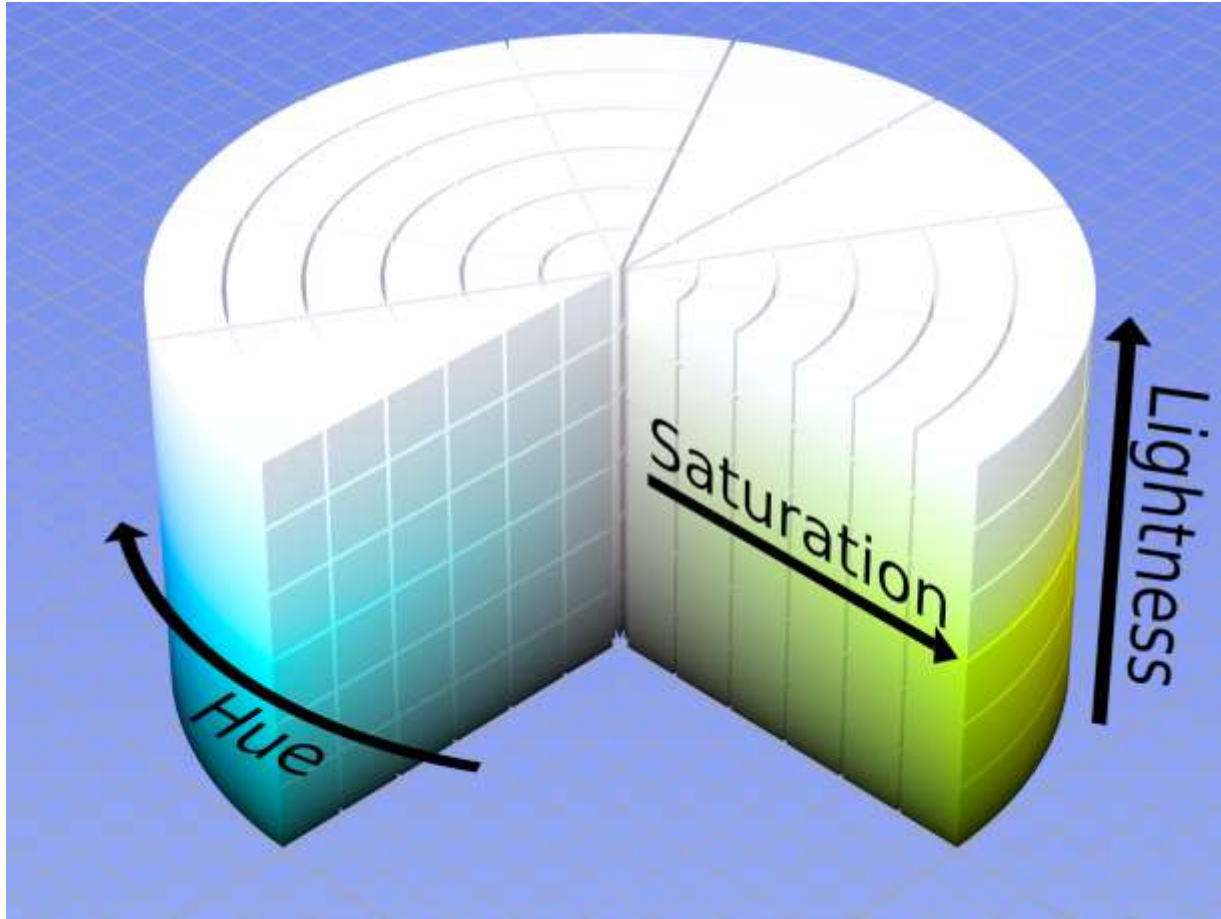
RGB versus CMYK



Model HSL [Wiki]

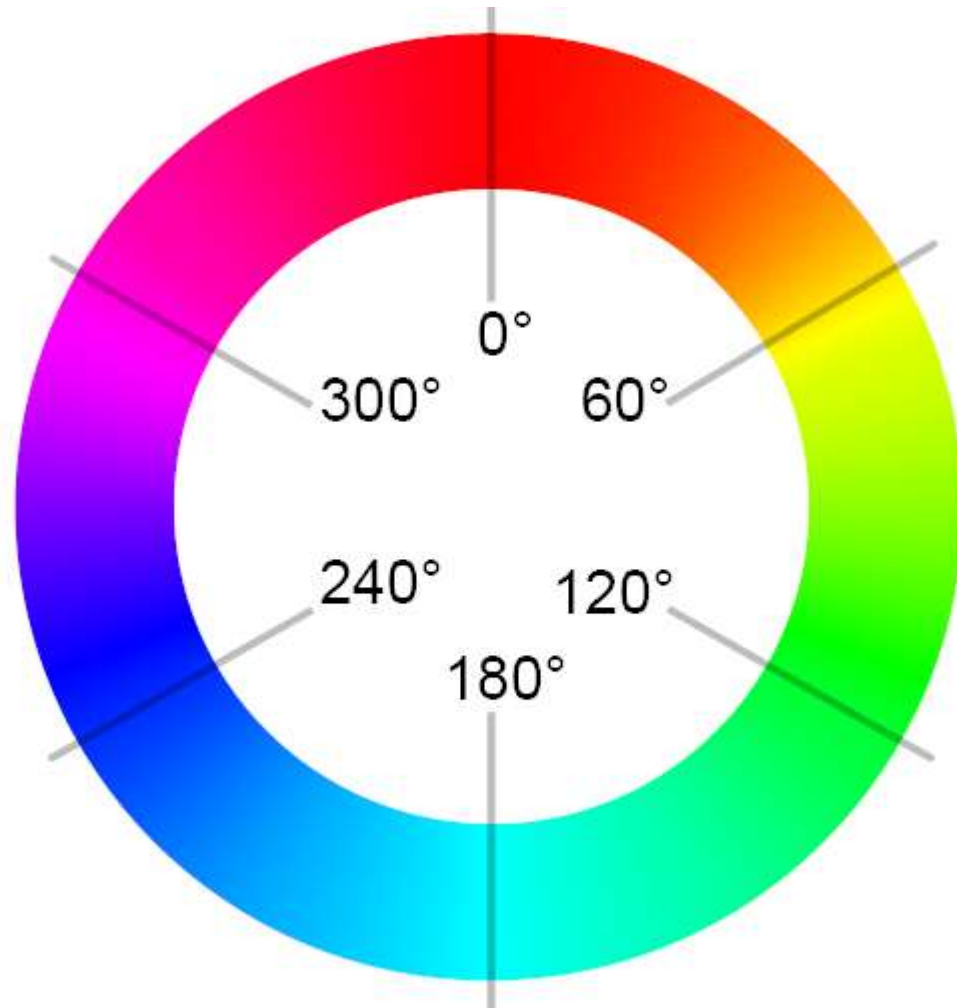
- Model z czasu startu telewizji (1926-30)
- **H: Hue** – (z ang. odcień, barwa),
 - o wartościach z przedziału: od 0 do 360 stopni.
- **S: Saturation** – nasycenie koloru.
 - z przedziału 0...1 albo 0...100%.
- **L: Lightness** – średnie światło białe,
 - z przedziału 0...1 albo 0...100%.

HSL



https://cs.pwr.edu.pl/cichon/2015_16_b/WW/W/img/HSL_color_solid_cylinder.png

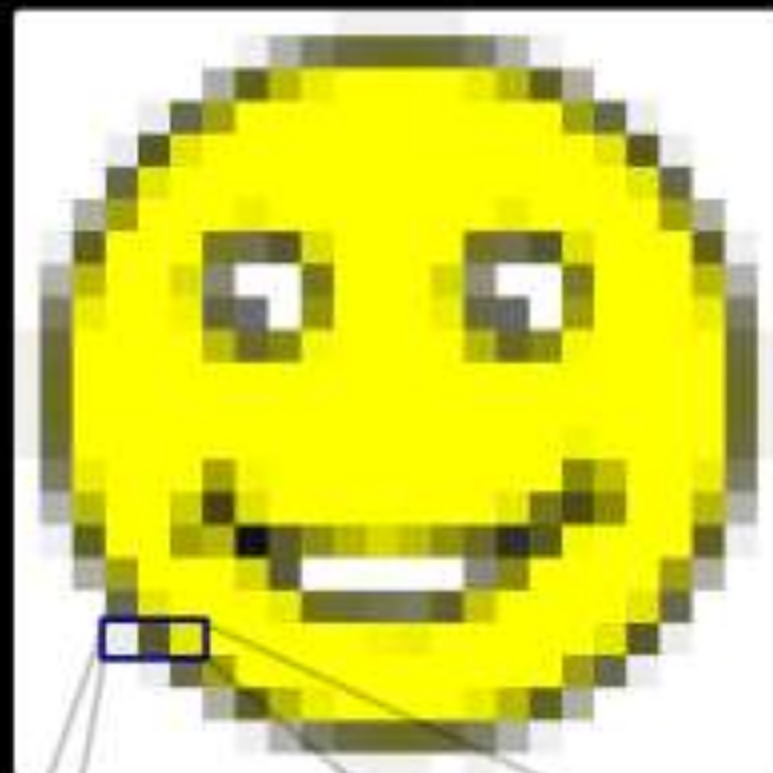
HSL - 360°



Grafika rastrowa [Wiki]

- Grafika rastrowa – prezentacja obrazu za pomocą pionowo-poziomej siatki odpowiednio kolorowanych pikseli na monitorze komputera, drukarce lub innym urządzeniu wyjściowym.
- Grafika rastrowa została opatentowana pierwszy raz przez firmę **Texas Instruments** w latach 70. i aktualnie jest wszechobecną formą przedstawiania obrazu cyfrowego.
- Zdjęcie z aparatu cyfrowego jest właśnie grafiką typu rastrowego.

Raster



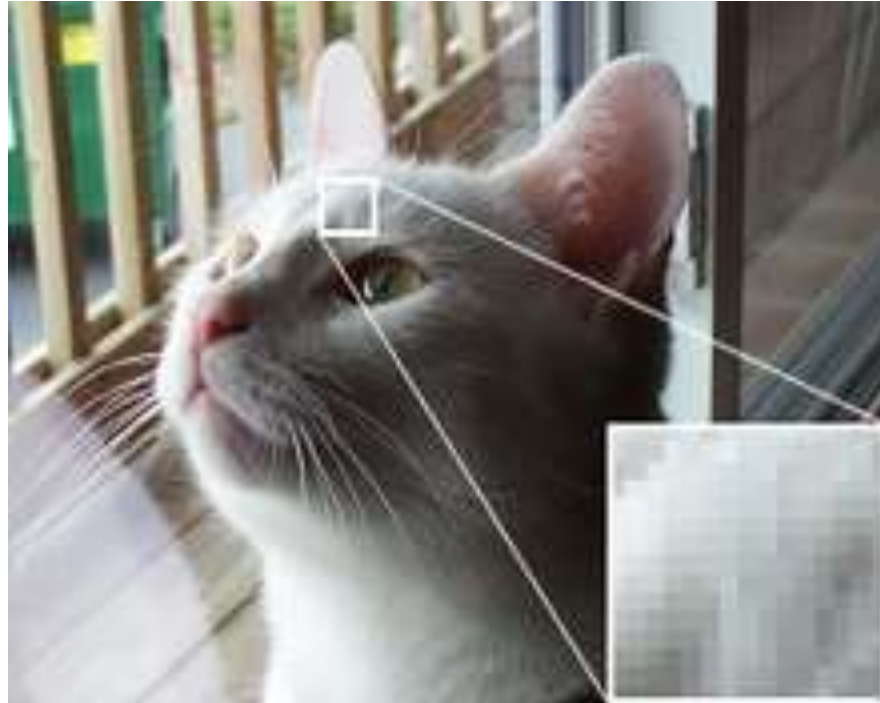
<https://upload.wikimedia.org/wikipedia/commons/thumb/3/3b/Rgb-raster-image.svg/300px-Rgb-raster-image.svg.png>

R 93%	R 35%	R 90%
G 93%	G 35%	G 90%
B 93%	B 16%	B 0%

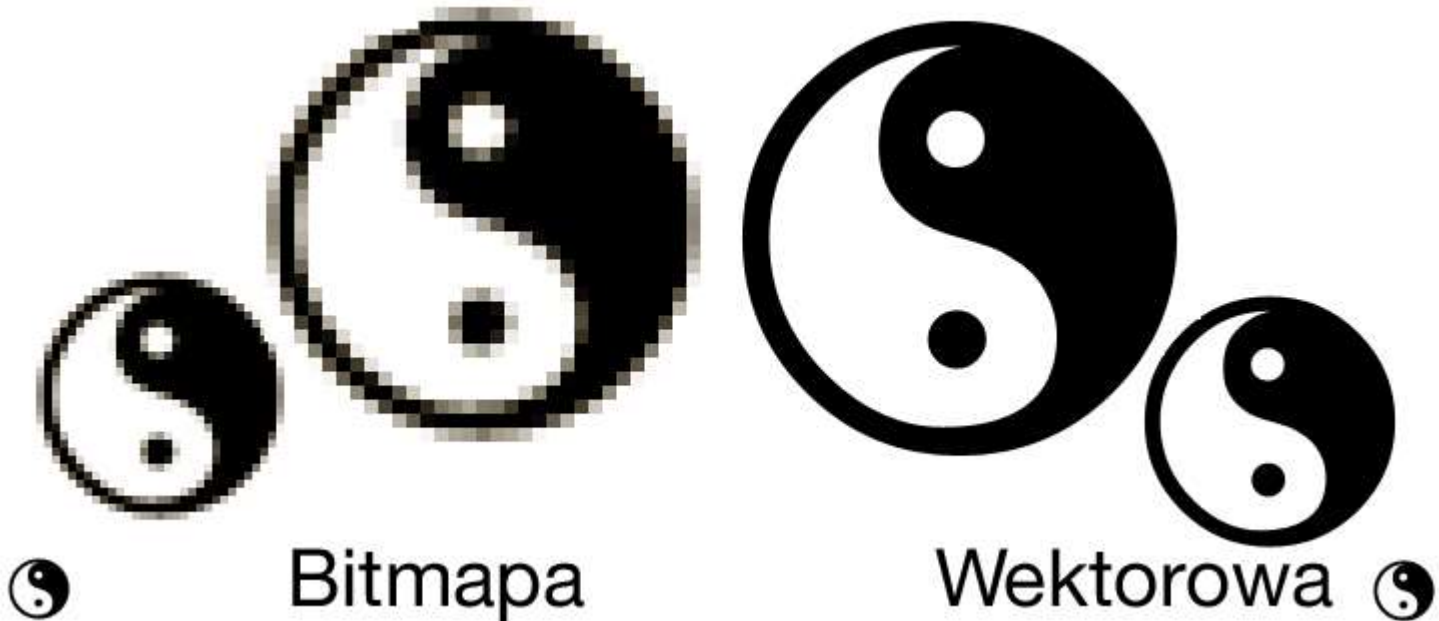
Pikselizacja [Wiki]

- **Pikselizacja** (potocznie: **pikseloza**) – zjawisko deformacji obrazu polegające na jego wyświetlaniu w sposób widoczny, ukazując poszczególne wiązki pikseli, które to nie są widoczne, gdy obraz jest widziany w pomniejszeniu, z większej odległości lub gdy zastosowano mniejszy stopień kompresji.
- W znacznym stopniu zjawisko można zredukować dzięki technikom redukcji artefaktów

Przykład pikselizacji



Problemy z bitmapami...



<http://roznice.com/wp-content/uploads/2013/11/bitmapa-wektor.jpg>

Grafika wektorowa [Wiki]

- Jeden z dwóch podstawowych rodzajów grafiki komputerowej,
- Obraz opisany jest za pomocą figur geometrycznych lub brył geometrycznych, umiejscowionych w matematycznie zdefiniowanym układzie współrzędnych
- Grafika wektorowa jest grafiką w pełni skalowalną
 - obrazy wektorowe można nieograniczenie powiększać oraz zmieniać ich proporcje bez uszczerbku na jakości.

Grafika obiektowa [Wiki]

- Nazwa **grafika obiektowa** – związana jest z faktem, iż obraz opisany jest za pomocą tzw. **obiektów**, które zbudowane są z podstawowych elementów nazywanych **prymitywami**, czyli
 - prostych figur geometrycznych takich jak odcinki, krzywe, okręgi, wielokąty.
- Każdy z prymitywów opisywany jest za pomocą **parametrów**, np.
 - w przypadku odcinka – współrzędnych jego końców, a w przypadku okręgu – współrzędnych środka i długości promienia.

Atrybuty obiektów [Wiki]

- Obiekty takie mają także określone atrybuty mówiące np. o
 - grubości i kolorze linii,
 - kolorze wypełnienia figury lub
 - wypełnieniu niejednolitym jak wypełnienie gradientem albo wzorem, stopniu przezroczystości.
- Atrybuty zależą głównie od stosowanego standardu opisu grafiki wektorowej.

Czcionki TrueType [Wiki]

- TrueType – format fontów wymyślony przez firmę Apple jako antidotum na PostScript (za który trzeba było płacić).
- Stosowany na komputerach Macintosh od 1991 obecnie rozpowszechniony na wszystkich platformach.
- Znaki w TrueType opisane są za pomocą krzywych Béziera tylko drugiego stopnia, jednak jest to format znacznie bardziej skomplikowany technicznie.

PostScript [Wiki]

- PostScript – uniwersalny język opisu strony opracowany przez firmę Adobe Systems, będący obecnie standardem w zastosowaniach poligraficznych.
- PostScript różni się od innych języków programowania tym, że został specjalnie zaprojektowany do tworzenia znaków tekstowych i obrazów graficznych na drukowanych stronach.

.BMP [Wiki]

- Format pliku z grafiką bitmapową.
 - Opracowany pierwotnie dla systemu OS/2,
 - wykorzystywany później także w interfejsach systemów z rodziny Microsoft Windows™
 - jako wolny od patentów jest dostępny i – mimo dużych rozmiarów – popularny jako format przechowywania danych również na wszystkich pozostałych platformach.
- Linie obrazu zapisywane są od dołu do góry.
- W obrazach TrueColor punkty zapisywane są w kolejności B, G, R (trzy bajty),
- W obrazach o mniejszej liczbie kolorów zapisywane są one na mniejszej ilości bajtów – przykładowo 16-bitowa bitmapa (High Color) zapisuje na dwóch bajtach

Kolorki... [Wiki]

- **High color** – sposób wyświetlania przestrzeni barw na ekranie wyświetlacza o 15-bitowej lub 16-bitowej głębi kolorów.
 - 15-bitowa głębia kolorów umożliwia wyświetlenie 2^{15} , czyli 32 768 kolorów.
 - 16-bitowa głębia kolorów umożliwia wyświetlenie 2^{16} , czyli 65 536 kolorów, zakodowanych w przestrzeni RGB
- **True color** – sposób wyświetlania barw na ekranie monitora o 24-bitowej głębi kolorów, umożliwiający uzyskanie 16 777 216 kolorów.
 - W trybie true color każdy z kolorów podstawowych RGB jest kodowany 8-bitowo, czyli dla każdego z nich jest określonych 256 poziomów jasności.

.GIF [Wiki]

- **Graphics Interchange Format** –
 - format pliku graficznego z **kompresją bezstratną** stworzony w 1987 przez firmę CompuServe.
- W jednym bloku obrazu (najczęściej obraz składa się z jednego bloku) format GIF obsługuje 256 kolorów
 - Stosowany do grafik o niewielkiej liczbie kolorów i ostrych krawędziach
 - Raczej nie nadaje się do fotografii
 - Problemy z **algorytmem LZW** do kompresji

Kompresja bezstratna [Wiki]



- **Kompresja bezstratna (lossless compression)**
 - ogólna nazwa metod kompresji informacji do postaci zawierającej zmniejszoną liczbę bitów, pod warunkiem, że metoda ta gwarantuje możliwość odtworzenia informacji z postaci skompresowanej do identycznej postaci pierwotnej.
- Algorytmy kompresji bezstratnej dobrze kompresują "typowe" dane, czyli takie w których występuje znaczna nadmiarowość informacji (**redundancja**).

.JPG [Wiki]

- Joint Photographic Experts Group) – algorytm **stratnej kompresji grafiki rastrowej**, wykorzystany w formacie plików graficznych o tej samej nazwie.
- Prace nad standardem rozpoczęły się w kwietniu 1983 roku w organizacji ISO.
- Standard został opublikowany w dwóch częściach w 1991 roku i definiował podstawowy sekwencyjny tryb kompresji stratnej, oparty na **dyskretnej transformacie kosinusowej (DCT)**

Kompresja stratna [Wiki]

- Metoda zmniejszania liczby bitów potrzebnych do wyrażenia danej informacji, które nie dają gwarancji, że odtworzona informacja będzie identyczna z oryginałem.
 - Dla niektórych danych algorytm kompresji stratnej może odtworzyć informację w sposób identyczny.
 - Jest możliwa ze względu na sposób działania ludzkich zmysłów, tj. wyższą wartość pewnych zakresów danych nad innymi.

Kompresja stratna [Wiki]

- Algorytmy kompresji stratnej zazwyczaj posługują się modelami psychoakustycznymi, psychowizualnymi itd., aby odrzucić najmniej istotne dane o dźwięku, obrazie, pozostawiając dane o wyższej wartości dla rozpoznawania tej informacji (akustycznej, wizualnej) przez zmysły.
- Ilość odrzucanych danych jest zazwyczaj określana przez **stopień kompresji**.

.PNG [Wiki]

- **Portable Network Graphics** –
 - Rastrowy format plików graficznych oraz system bezstratnej kompresji danych graficznych.
 - PNG został opracowany jako następca GIF w 1995 roku po ogłoszeniu roszczeń patentowych dotyczących kompresji LZW używanej w GIF.
 - Przewiduje tryby o niskiej ilości kolorów (paletowe od 2 do 256 oraz tryb skali szarości), przydatne w sytuacjach, gdy nadmiarowa informacja jest niewskazana (tj. gdy ważniejsza od subiektywnej wierności oryginałowi jest treść obrazu).
 - Ma także tryb 48-bitowego RGB,

.SVG

- **Scalable Vector Graphics –**

- uniwersalny format dwuwymiarowej grafiki wektorowej (statycznej i animowanej), nieobwarowany licencjami i patentami.
- SVG został stworzony w 1999 przez jedną z najważniejszych organizacji zajmujących się standardami w Internecie – W3C
- W SVG oprócz standardowych obiektów (prostokąty, elipsy, krzywe) można opisywać efekty specjalne (filtry), maski przezroczystości, wypełnienia gradientowe itp.
- SVG należy do rodziny XML

Cyfryzacja czyli digitalizacja