



유니코드 알아보기



아스키코드 부터 유니코드까지



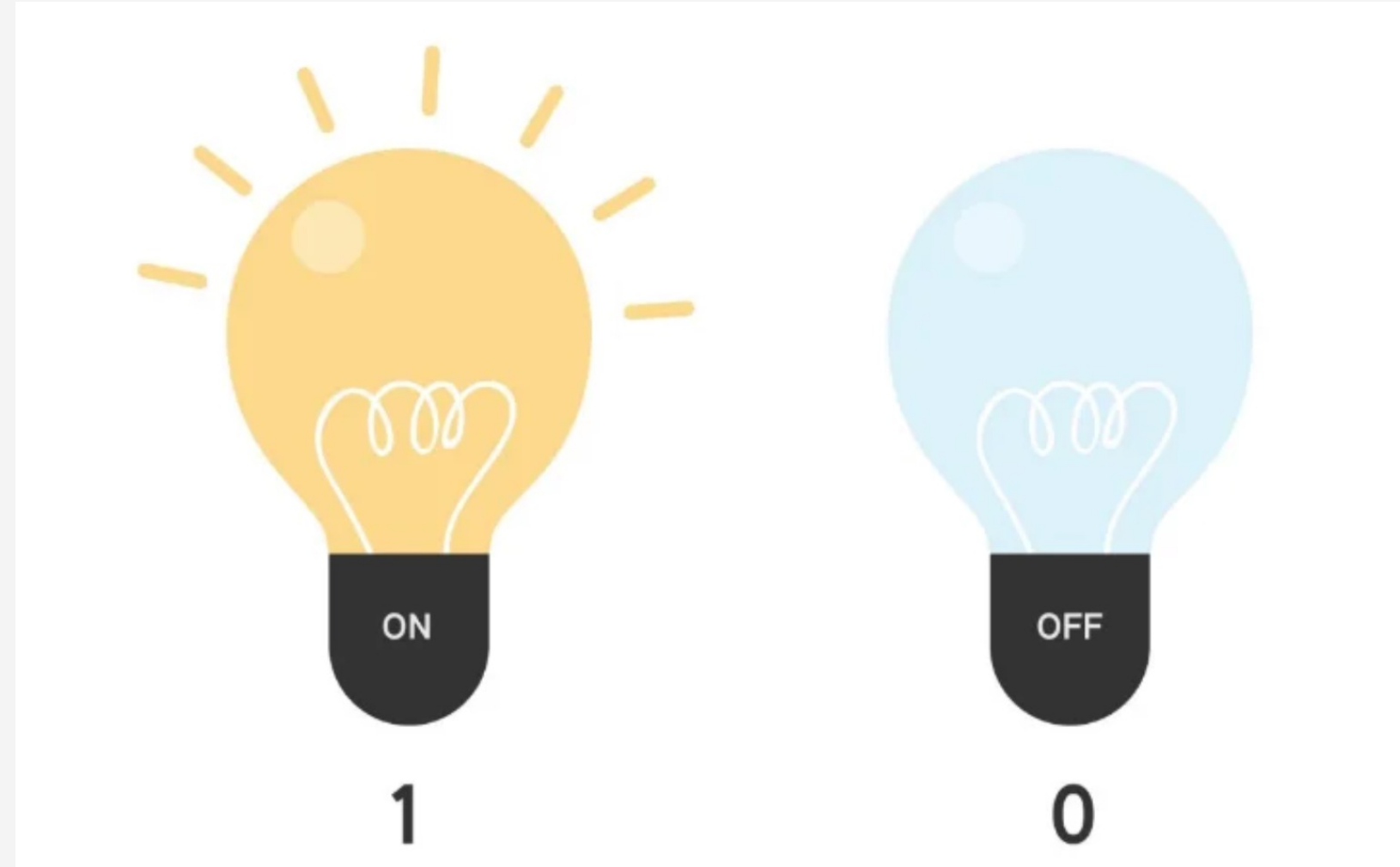
유니코드 인코딩

Q&A

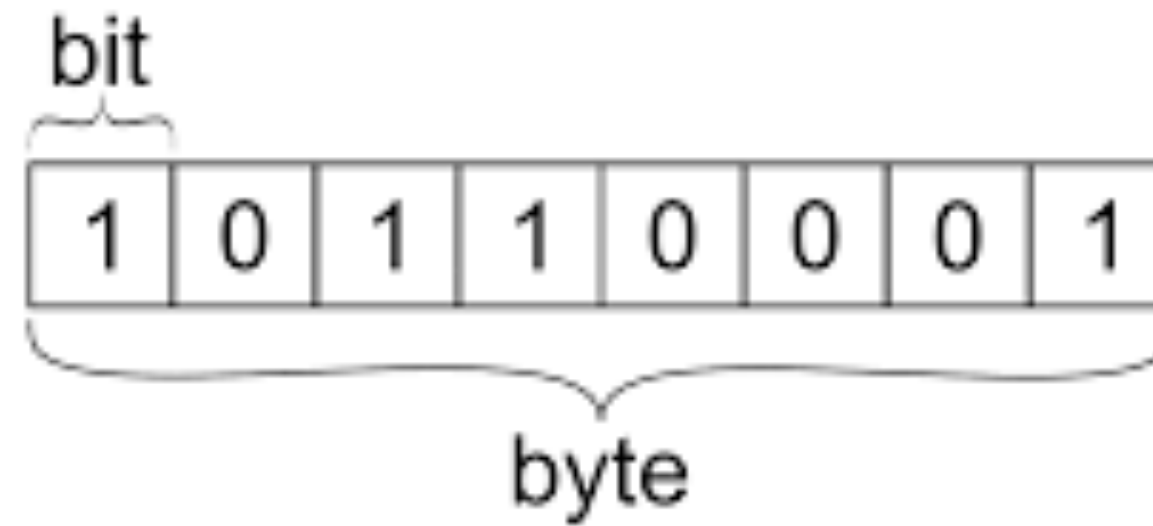
Q&A



컴퓨터의 탄생



전기신호가 있으면 '1'
전기신호가 없으면 '0'



8bit = 1byte

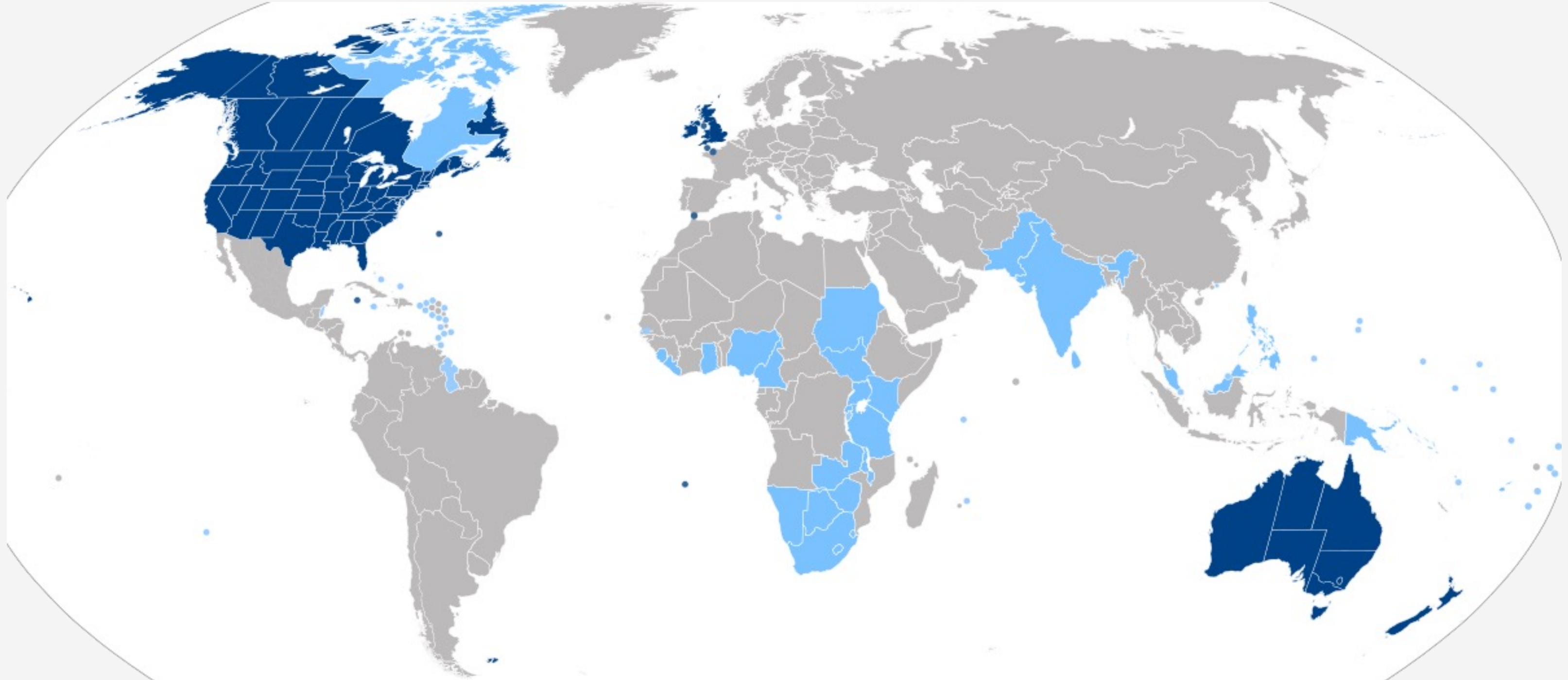


0과 1밖에 모르는 컴퓨터

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

Source: www.asciitable.com

아스키 코드



영어권 문화



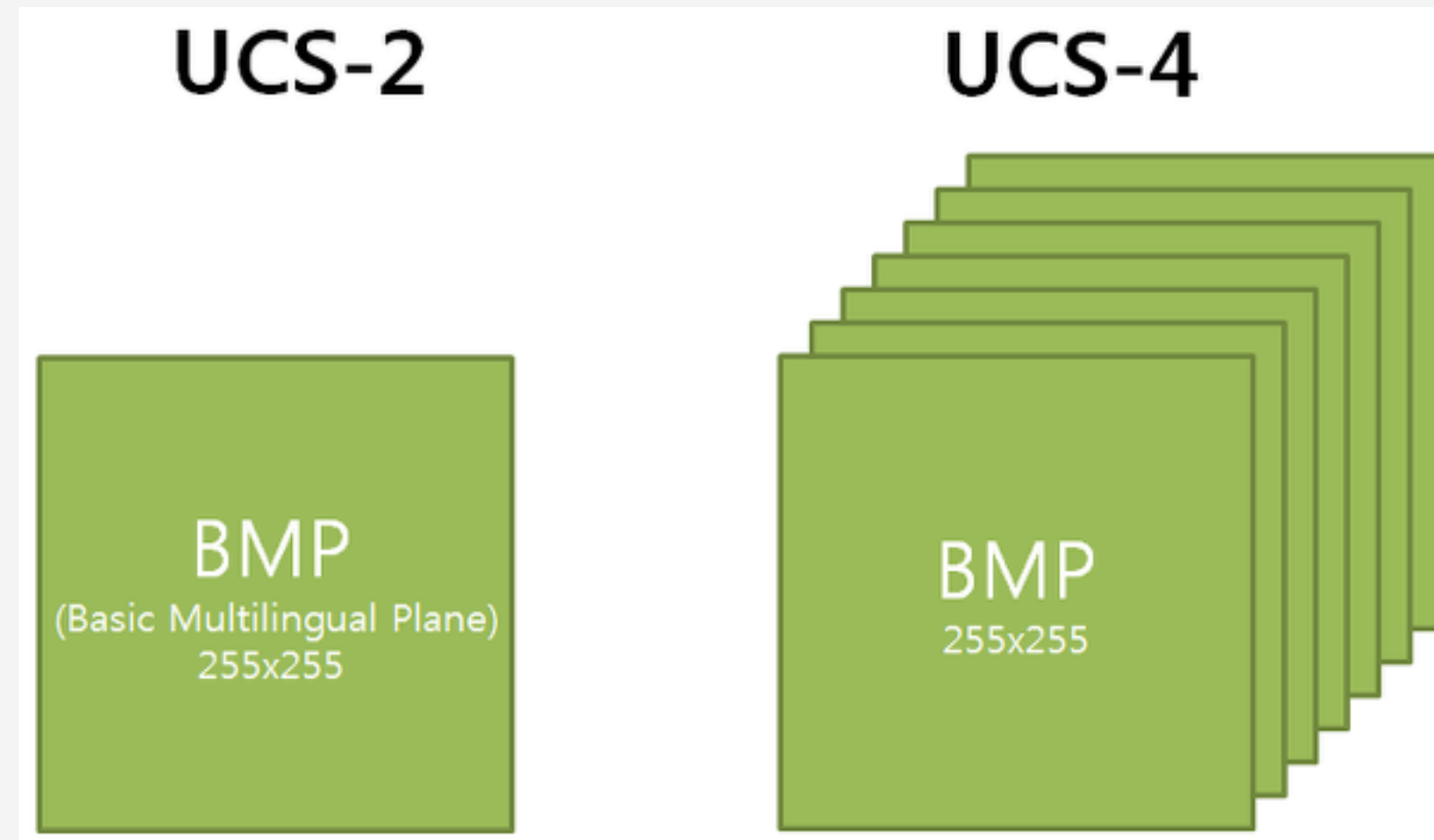
전세계 언어에 대해 적용



유니코드 등장!



2바이트 = 16비트 = 2^{16} = 65536



2바이트 > > 4바이트

UCS-2



2바이트



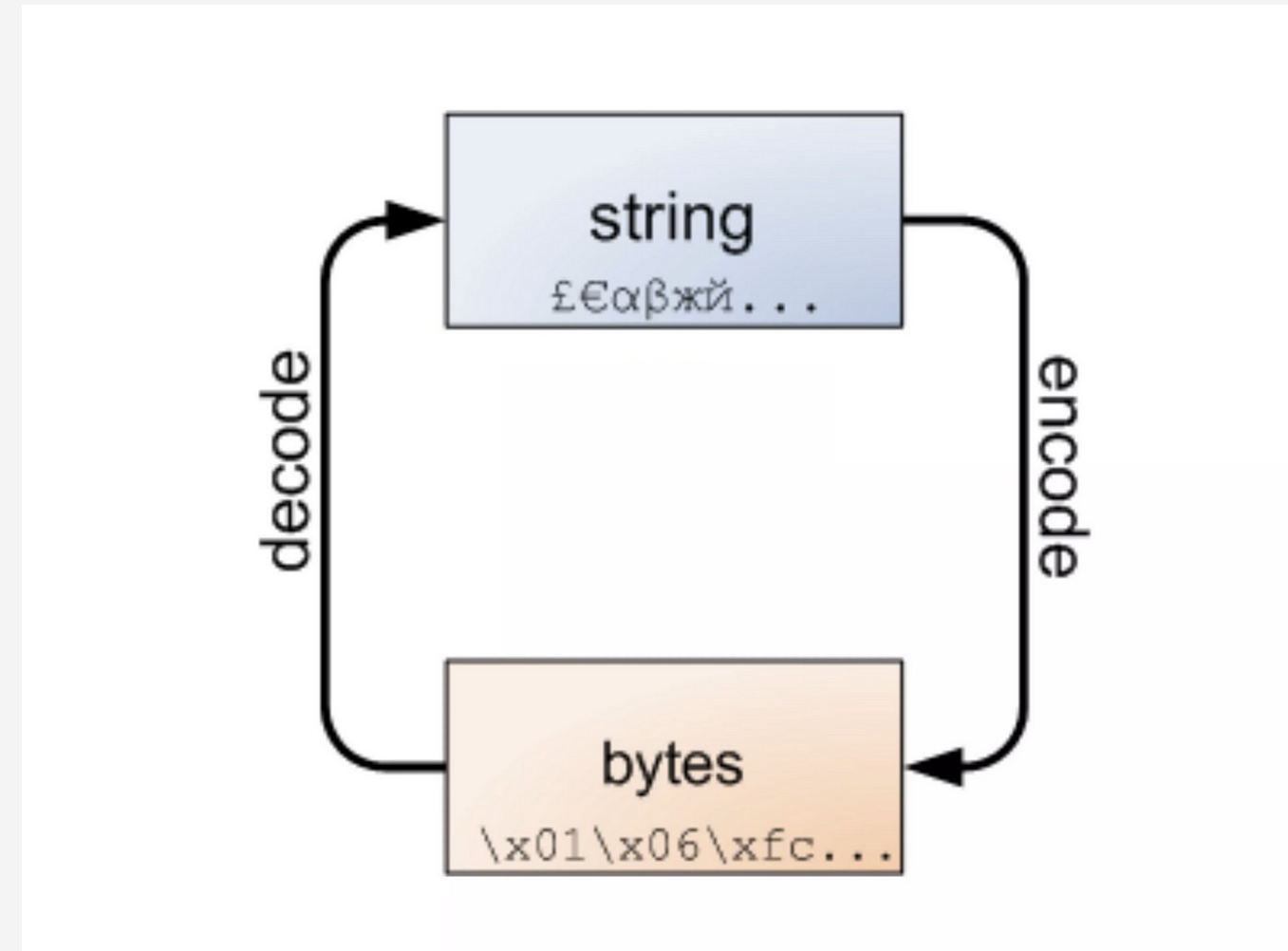
1 바이트



2 바이트



3 바이트..



인코딩



UTF-8

Universal Transformation Format - 8bit



1 바이트



2 바이트

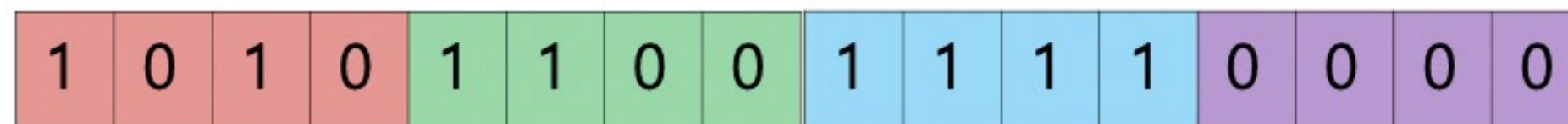


3 바이트..

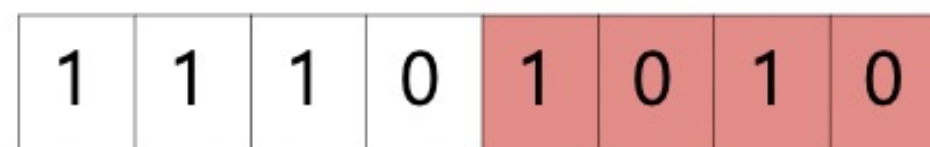
Unicode "곰" : U+ACF0



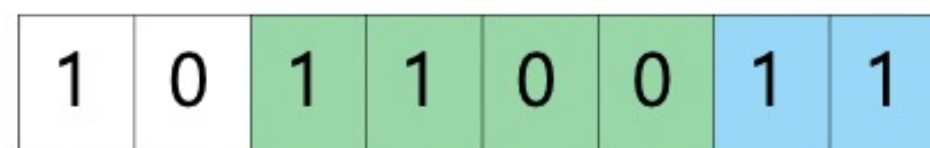
이진수 표현



UTF-8 인코딩



3byte 이기때문에 3byte 표식(1110)



나머지 byte 표식(10)

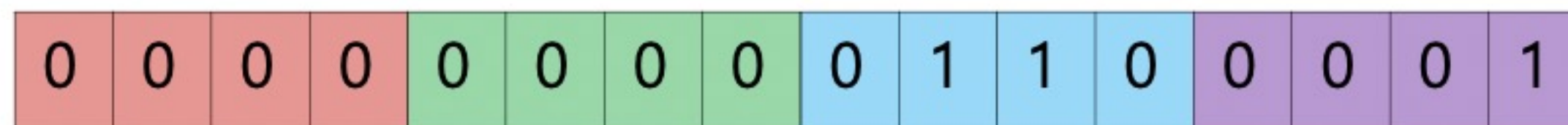


나머지 byte 표식(10)

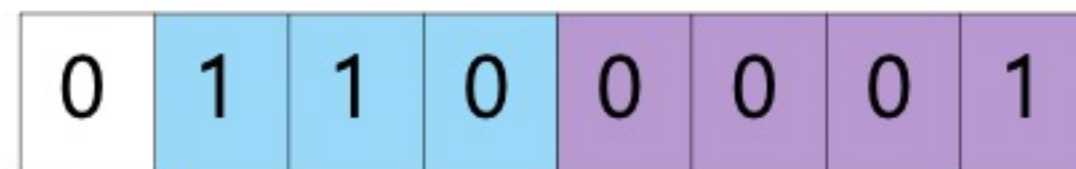
Unicode "a" : U+0061



이진수 표현



UTF-8 인코딩



1byte 이기때문에, 1byte표식(0)

UTF-8
1~4 byte

UTF-16
2~4 byte

00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F	로마자, 로마자권 기호
10	11	12	13	14	15	16	17	18	19	1A	1B	1C	1D	1E	1F	기타 유럽 문자
20	21	22	23	24	25	26	27	28	29	2A	2B	2C	2D	2E	2F	아프리카 문자
30	31	32	33	34	35	36	37	38	39	3A	3B	3C	3D	3E	3F	중동·서남아시아 문자
40	41	42	43	44	45	46	47	48	49	4A	4B	4C	4D	4E	4F	남부와 중앙 아시아 문자
50	51	52	53	54	55	56	57	58	59	5A	5B	5C	5D	5E	5F	동남아시아 문자
60	61	62	63	64	65	66	67	68	69	6A	6B	6C	6D	6E	6F	동아시아 문자
70	71	72	73	74	75	76	77	78	79	7A	7B	7C	7D	7E	7F	CJK 문자
80	81	82	83	84	85	86	87	88	89	8A	8B	8C	8D	8E	8F	인도네시아, 오세아니아 문자
90	91	92	93	94	95	96	97	98	99	9A	9B	9C	9D	9E	9F	북미 및 남미 문자
A0	A1	A2	A3	A4	A5	A6	A7	A8	A9	AA	AB	AC	AD	AE	AF	Notational systems
B0	B1	B2	B3	B4	B5	B6	B7	B8	B9	BA	BB	BC	BD	BE	BF	기호
C0	C1	C2	C3	C4	C5	C6	C7	C8	C9	CA	CB	CC	CD	CE	CF	사용자 정의 영역
D0	D1	D2	D3	D4	D5	D6	D7	D8	D9	DA	DB	DC	DD	DE	DF	UTF-16 상·하위 대체 영역
E0	E1	E2	E3	E4	E5	E6	E7	E8	E9	EA	EB	EC	ED	EE	EF	
F0	F1	F2	F3	F4	F5	F6	F7	F8	F9	FA	FB	FC	FD	FE	FF	

유니 코드 버전 15.1

UTF-16
2~4 byte

기본 다국어 평면

UTF-8
1~4 byte

VS

UTF-16
2~4 byte



Q & A