

文字のデジタル表現について

｜年情報

コンピュータで文字を表すには・・・

文字ごとに数値をわりあて、その数値で文字を区別する

①文字コード・・・文字にわりあてられる数値



図9 「CAT」という文字列のデジタル表現

日本工業規格 (JIS) で定められた JIS X0201 という文字コード

- 左半分は ①ASCIIコード (ASCII)

英字や数字を
表す

										SP…スペース DEL…文字消去							
下位の桁	上位の桁	0000 0	0001 1	0010 2	0011 3	0100 4	0101 5	0110 6	0111 7	1000 8	1001 9	1010 A	1011 B	1100 C	1101 D	1110 E	1111 F
0000	0	コンピュータを制御するために 使われる記号(省略)		SP	0	@	P	`	p	未定義	未定義	未定義	—	タ	ミ	未定義	未定義
0001	1			!	1	A	Q	a	q			。	ア	チ	ム		
0010	2			“	2	B	R	b	r			「	イ	ツ	メ		
0011	3			#	3	C	S	c	s			」	ウ	テ	モ		
0100	4			\$	4	D	T	d	t			、	エ	ト	ヤ		
0101	5			%	5	E	U	e	u			・	オ	ナ	ユ		
0110	6			&	6	F	V	f	v			ヲ	カ	ニ	ヨ		
0111	7			‘	7	G	W	g	w			ア	キ	ヌ	ラ		
1000	8			(8	H	X	h	x			イ	ク	ネ	リ		
1001	9)	9	I	Y	i	y			ウ	ケ	ノ	ル		
1010	A			*	:	J	Z	j	z			エ	コ	ハ	レ		
1011	B			+	;	K	[k	{			オ	サ	ヒ	ロ		
1100	C			,	<	L	¥	l				ヤ	シ	フ	ワ		
1101	D			—	=	M]	m	}			ユ	ス	ヘ	ン		
1110	E			.	>	N	^	n	—			ヨ	セ	ホ	°		
1111	F			/	?	O	—	o	DEL			ツ	ソ	マ	°		

- 右半分は
日本語のカタカナ
などを表す

表 6 JIS X0201文字コード

- 日本語にはカタカナ、ひらがな、漢字などがあり常用漢字だけでも2136文字ある。ASCIIコードでは7ビットで128通り（128文字）しか表現できない。
- そこで日本語の文字は1文字を16ビット（2バイト）で表現する。
16ビットで表現できるのは ①65535 通り（文字）までである。
文字コードとして ②Shift_JIS コード ③EUC が用いられる。

Unicode

世界中の文字を、1つの文字コードの枠組みの中で、
統一的に表すために設計された文字コード

コードのわりあて方のちがいにより、**UTF-8** や**UTF-16** などの方式がある

同じ文字でも文字コードの体系が異なれば、コードは異なる

文字コードの体系	文字コード(16進法)	
	日	本
JISコード	46 7C	4B 5C
Shift_JISコード	93 FA	96 7B
EUC	C6 FC	CB DC
Unicode(UTF-8)	E6 97 A5	E6 9C AC

異なる文字コードの体系で文字を表示させると・・・

本来の文字とは異なる文字が表示される



⑥文字化け の原因の1つ

馴!@著肴嫌□@著肴塘□闢 n
0□々演□著肴塘□・A 演劇的
w”・枉死俣・@u”縹艸奄劇
能的t紙・阨L l”・烙摩闊
l□々演遠□@俳優、舞踊遠□/
□・A 又w)縁o l”・凰閨÷

図10 文字化け

問題①、② | love→2進数、16進数

7

2進数 0110 1100 0110 1111
 0111 0110 0110 0101

16進数 6C 6F 76 65

問題③ | 浮かび上がる文字は？

8

0110 1011 0110 1111 0111 1010 0110 1111

答え K0Z0

①ひらがなや漢字を1文字表示するには何ビット必要ですか。

答え 16ビット (2バイト)

②①が全部で200文字だと何ビット必要ですか。。

200文字×16ビット

答え 3200ビット

① ②が2ページだとすると何ビット必要ですか。

$$3200 \text{ ビット} \times 2$$

答え 6400ビット

② ③をB（バイト）になおすとどうなりますか

$$6400 \div 8$$

答え 800バイト