

第1編 コンピュータとデジタル情報

第0章 位取記数法

第1章 コンピュータのしくみ

第2章 情報のデジタル表現

続(文字・音・画像・動画・圧縮等)

データのデジタル表現の方法

コンピュータはデータをどんな規則でデジタル表現するのか？ その方法を個別に学習する。

基本は
1対1対応

数値のデジタル表現	固定小数点表示(整数), 浮動小数点表示(実数)
文字のデジタル表現	文字コード, 1バイト文字, 2バイト文字, (フォント)
音声のデジタル表現	標本化, 量子化, 符号化, D-A変換, A-D変換
画像のデジタル表現	標本化, 量子化, 符号化, 解像度, 3原色, 静止画, 動画, ペイント系, ドロー系
データの容量と圧縮	圧縮, 可逆性, 非可逆性, 解凍(伸張)
ファイル形式と拡張子	品質, 特徴

文字のデジタル表現: 基本

関連: pp.32-33

コンピュータにおける文字の処理

文字を入力するとコンピュータは,

- ⇒ 規則に従って, **文字とビット列**(0,1の並び)を**1対1対応**させる。
- ⇒ データ保存は, **ビット列**(0,1の並び)の状態です。
- ⇒ 画面や印刷物に, **ビット列に対応した文字の形(フォント)**を表示させる。

基本は
1対1対応

よく使われるビット列の長さは**7, 8, 16** (24,32,...もあり)

1バイト文字 英字や数字やよく使う記号は7ビット(128種類)で十分。
(半角文字) カナを使うには1バイト必要(半角カタカナを追加!)

2バイト文字 日本語をすべて使用するには2バイト(65,536種類)で安心。
(全角文字) (重複するが英字・数字・記号・カタカナも含めて)

古くから使われている文字集合

関連: pp.32-33

ASCII

JIS X 0201

	0000	0001	0010	0011	0100	0101	0110	0111	1000	1001	1010	1011	1100	1101	1110	1111
0000	00	NUL	SOH	STX	ETX	END	ACK	BEL	BS	HT	LF	VT	FF	CR	DEL	
0001	10	DLA	DC1	DC2	DC3	DC4	NAK	ETB	CAN	EM	ENQ	ESC	FR	GR	HR	
0010	20		!	"	#	\$	%	&	'	()	*	+	,	-	.
0011	30	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>
0100	40	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
0101	50	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^
0110	60	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n
0111	70	p	q	r	s	t	u	v	w	x	y	z	{		}	~

これだけなら
8bit(1Byte)で
足ります。
(7bitで十分)

問い: 表を用いて **Kobe** を, 16進数と2進数で表そう(ノートに)。

問い: 表を用いて
Kobe を,
16進数と
2進数で表そう。
答:

K 4B 0100 1011
o 6F 0110 1111
b 62 0110 0010
e 65 0110 0101



古くから使われている文字集合

関連: pp.32-33

ASCII

JIS X 0201

	0000	0001	0010	0011	0100	0101	0110	0111	1000	1001	1010	1011	1100	1101	1110	1111
0000	00	NUL	SOH	STX	ETX	END	ACK	BEL	BS	HT	LF	VT	FF	CR	DEL	
0001	10	DLA	DC1	DC2	DC3	DC4	NAK	ETB	CAN	EM	ENQ	ESC	FR	GR	HR	
0010	20		!	"	#	\$	%	&	'	()	*	+	,	-	.
0011	30	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>
0100	40	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
0101	50	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^
0110	60	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n
0111	70	p	q	r	s	t	u	v	w	x	y	z	{		}	~
1000	80															
1001	90															
1010	A0															
1011	B0															
1100	C0															
1101	D0															
1110	E0															
1111	F0															

これだけなら
8bit(1Byte)で
足ります。
(7bitで十分)



補足: ISO646 (現ISO/IEC646)に基づく文字割当例

コード番号	0x23	0x24	0x40	0x5B	0x5C	0x5D	0x5E	0x60	0x7B	0x7C	0x7D	0x7E
ASCII	#	\$	@	[\]	^	`	{		}	~
日本	#	\$	@	[¥]	^	`	{		}	~
イギリス	£	\$	@	[\]	^	`	{		}	~
ドイツ	#	\$	§	Ä	Ö	Ü	^	`	ä	ö	ü	ß
フランス	£	\$	à	°	ç	§	^	μ	é	ù	è	~

文字のデジタル表現: 用語の定義

関連: pp.32-33

よく使われる専門用語の定義

- ① **文字集合** (character set) :
文字の一覧。**符号化文字集合** (coded character set) 等ともいう。
- ② **文字符号化方式** (character encoding scheme) :
文字集合①に含まれる各文字に1対1で対応つけた記号を、データ列(**ビット列**)に変換する方法(規則)。
- ③ **文字コード** (character code) :
文字符号化方式②で各文字に対応したビット列。
- ④ **文字コード体系** (character encoding system) :
文字コードの一覧。単に**文字コード**と呼ぶことも多い(紛らわしい)。
- ⑤ **エンコーディング** (符号化処理) :
文字等のデータをビット列に置き換える処理。
逆に元に戻すことを、**デコーディング**という。

補足: 紛らわしい専門用語の解説

関連: pp.32-33

- ①文字集合 (**符号化文字集合**) ③文字コード
- ②文字符号化方式 ④文字コード体系, 等

A:人間用の符号	B:文字例
1-4-73	ら
1-4-74	り
1-4-75	る

- JIS規格等, 「A:とB:の対応表」が複数種類存在する。
「①**文字集合**」とは, これら「A:とB:の対応表」のこと。
- コンピュータは数値(ビット列)しか扱えないから「B:とC:の対応表」も必要。
「④**文字コード体系**」とは, この「B:とC:の対応表」のこと。
- ④を作る際の(すなわち数値化するための)規則が②であり, ②の規則に沿ってできた文字ごとの数値(ビット列)が③である。

補足: 紛らわしい専門用語の解説

関連: pp.32-33

- ①文字集合 (符号化文字集合) ③文字コード
- ②文字符号化方式 ④文字コード体系, 等

B:文字例	C:コンピュータ用符号 (=数値)
ら	1000 0010 1110 0111 (82E7)
り	1000 0010 1110 1000 (82E8)
る	1000 0010 1110 1001 (82E9)

- JIS規格等, 「A:とB:の対応表」が複数種類存在する。
「①**文字集合**」とは, これら「A:とB:の対応表」のこと。
- コンピュータは数値(ビット列)しか扱えないから「B:とC:の対応表」も必要。
「④**文字コード体系**」とは, この「B:とC:の対応表」のこと。
- ④を作る際の(すなわち数値化するための)規則が②であり, ②の規則に沿ってできた文字ごとの数値(ビット列)が③である。

補足: 紛らわしい専門用語の解説

関連: pp.32-33

- ①文字集合 (**符号化文字集合**) ③文字コード
- ②文字符号化方式 ④文字コード体系, 等

A:人間用の符号	B:文字例	C:コンピュータ用符号 (=数値)
1-4-73	ら	1000 0010 1110 0111 (82E7)
1-4-74	り	1000 0010 1110 1000 (82E8)
1-4-75	る	1000 0010 1110 1001 (82E9)

- JIS規格等, 「A:とB:の対応表」が複数種類存在する。
「①**文字集合**」とは, これら「A:とB:の対応表」のこと。
- コンピュータは数値(ビット列)しか扱えないから「B:とC:の対応表」も必要。
「④**文字コード体系**」とは, この「B:とC:の対応表」のこと。
- ④を作る際の(すなわち数値化するための)規則が②であり, ②の規則に沿ってできた文字ごとの数値(ビット列)が③である。

文字のデジタル表現: 用語の定義

関連: pp.32-33

よく使われる専門用語の定義

- ① **文字集合** (character set) :
文字の一覧。**符号化文字集合** (coded character set) 等ともいう。
- ② **文字符号化方式** (character encoding scheme) :
文字集合①に含まれる各文字に1対1で対応つけた記号を、データ列(**ビット列**)に変換する方法(規則)。
- ③ **文字コード** (character code) :
文字符号化方式②で各文字に対応したビット列。
- ④ **文字コード体系** (character encoding system) :
文字コードの一覧。単に**文字コード**と呼ぶことも多い(紛らわしい)。
- ⑤ **エンコーディング** (符号化処理) :
文字等のデータをビット列に置き換える処理。
逆に元に戻すことを、**デコーディング**という。

文字コード体系の例

関連: pp.32-33

(符号化)文字集合 (Coded) character set	文字符号化方式 character encoding scheme	文字コード(体)系, 文字コード character encoding system, character code
ASCII	なし	ASCII 7bit
JIS X 0201 & JIS X 0208 等	ISO-2022-JP Shift_JIS	JIS(漢字)コード 1,2Byte(7bit表現) Shift_JISコード 1,2Byte
ISO-2022-JP: 半角の片仮名含まず	EUC	EUC-JPコード 1,2Byte
Unicode	UTF-8, UTF-8N	UTF-8 ASCII部分: 1Byte 他: 2-6Byte Web で使用
	UTF-16BE, UTF-16LE	UTF-16 2Byte or 4Byte
	UTF-32BE, UTF-32LE	UTF-32 4Byte固定長

文字集合の例(変更の頻度は著しい)			
名称・規格番号	通称・俗称	文字総数	備考
ASCII		128	1963年制定
JIS X 0201	ANKコード	159	1969年制定
JIS X 0208	JIS78(旧JIS)	6802	1978年
	JIS83(新JIS)	6877	1983年
	JIS90	6879	1990年
JIS X 0212	97JIS	6879	1997年 第1水準・第2水準
	補助漢字	6067	1990年
JIS X 0213	JIS2000	11223	2000年 第3水準・第4水準
	JIS2004	11233	2004年
Unicode		136690	1991年制定, 頻繁に改良

関連: pp.32-33

整理: 文字のデジタル表現の具体例

関連: pp.32-33

1バイト文字 英字や数字やよく使う記号は1バイト(256種類)で十分。
(半角文字) カタカナも含めてもOK(このカタカナを半角カタカナという)。

2バイト文字 日本語をすべて表示するには2バイト(65,536種類)欲しい。
(全角文字) (全角文字の英字・数字・記号・カタカナも作られた)

知っておくべき文字コード体系の種類

- ① ASCII (情報交換用米国標準コード1963) 7ビット
- ② JIS (ISO-2022-JP メールで使われる) 8,16ビット(1.2バイト)
- ③ Shift_JIS (Microsoft社, Apple社が使用) 8,16ビット(1.2バイト)
- ④ EUC-JP (Extended Unix Code: UNIX, Linuxで使用) 8,16ビット(1.2バイト)
- ⑤ UTF-** (世界各国の言語に対応) 3バイト以上使用される場合有



参考: IMEパッド 文字一覧(Shift_JISの一部)

補足: 文字化け・・・3つの原因

関連: pp.32-33

- 半角カタカナの使用
- 機種依存文字の使用
- 文字コード系 指定間違い

代表的な機種依存文字一覧
<https://meot-japan.com/2019/guide/izon.html>

キャリア6社の絵文字が統一
http://k-tai.impress.co.jp/docs/news/20140424_646054.html

絵文字
<http://www.softbank.jp/mobile/service/mail/3g/pictgram/>

参考: 文字のデザイン(表示・印刷)

関連: pp.-

グリフ 各文字の形(デザイン) **フォント** グリフの一覧

等幅フォント 文字幅が均一化(統一)されている

プロポーションアルフォント 文字によって文字幅が異なる

ビットマップフォント 点の配置関係でグリフを決定する

アウトラインフォント 輪郭線を数式にしてその都度計算してグリフを決定する

セリフ **サンセリフ** 文字の装飾による分類

pt(ポイント) 文字サイズの単位 1 pt = 約0.35mm

http://culture.cc.hirosaki-u.ac.jp/english/utsumi/info/moji_c6_ja.html