文字コード

コンピュータが直接扱えるのは、0 と 1 で表わされる 2 進数だけです。一方、人間は、数値以外のデータである文字、音、色などもコンピュータで取り扱いたいわけです。そこで、文字、音、色などの数値でないデータを 2 進数の数値を使って表すために、符号化（以下、コード化）という工夫が必要になります。

コード化は、コンピュータを使う人間が自由勝手に行うこともできますが、それだと、機種や端末装置が変わったり、インターネット等を介して相互に情報をやりとりすると、正しく表示されない（このことを一般に「化（ば）ける」といいます）事態になります。
ですので、一般的には、コード化は予め決められた符号化のルール（文字コード体系）に従って行います。

ここでは、文字（文字の中には 10 進数の数字も含まれます）のためのコード化について解説します。

コンピュータの世界では統一的な文字コード体系が定められていますが、残念なことに、1 種類ではなく、代表的なものだけでも数種類あります。皆さんの中で、パソコンを利用している際に、「文字化け」を経験したことはありませんか。これは、文字コード体系が複数あるのが原因なのです。

以下、代表的なものを紹介します。
なお、コンピュータでは 2 進数ですが、それだと長くなり、また読みにくいので、16 進数で表記します。

ASCII

ASCII（American Standard Code for Information Interchange の略で、アスキーと呼びます）は、米国で生まれた文字コード体系で、文字コード体系の元祖です。

英数記号の 1 文字を 7 ビットで表すもので、大文字・小文字の英字、数字、記号といった文字が 95 個（内、1 個は空白文字）と、33 個の制御文字の計 128 文字が規定されています。

制御文字は、表示するための文字種ではなく、モニタやプリンタなどの入出力装置を制御するためのものです。

16進数	ASCII	16進数	ASCII	16進数	ASCII	16進数	ASCII
00	NULL	20	SP	40	@	60	`
01	SOH	21	!	41	A	61	a
02	STX	22	"	42	B	62	b
03	ETX	23	#	43	C	63	c
04	EOT	24	$	44	D	64	d
05	ENG	25	%	45	E	65	e
06	ACK	26	&	46	F	66	f
07	BEL	27	'	47	G	67	g
08	BS	28	(	48	H	68	h
09	HT	29	)	49	I	69	i
0A	LF	2A	*	4A	J	6A	j
0B	VT	2B	+	4B	K	6B	k
0C	FF	2C	,	4C	L	6C	l
0D	CR	2D	-	4D	M	6D	m
0E	SO	2E	.	4E	N	6E	n
0F	SI	2F	/	4F	O	6F	o
10	DLE	30	0	50	P	70	p
11	DC1	31	1	51	Q	71	q
12	DC2	32	2	52	R	72	r
13	DC3	33	3	53	S	73	s
14	DC4	34	4	54	T	74	t
15	NAK	35	5	55	U	75	u
16	SYN	36	6	56	V	76	v
17	ETB	37	7	57	W	77	w
18	CAN	38	8	58	X	78	x
19	EM	39	9	59	Y	79	y
1A	SUB	3A	:	5A	Z	7A	z
1B	ESC	3B	;	5B	[	7B	{
1C	FS	3C	<	5C	\	7C	|
1D	GS	3D	=	5D	]	7D	}
1E	RS	3E	>	5E	^	7E	~
1F	US	3F	?	5F	_	7F	DEL

※赤は制御文字です（印字されません）
※SP（16進数の20）は半角の空白（ブランク）
※16進数の5Cは、Windows OSの和文フォントでは¥で表示されます

ASCII コード表

【便利知識】
1 バイトは 8 ビットですが、英字・数字・記号を扱うには 7 ビットあれば十分なので、データ長を短くしてデータ通信の負荷を軽減するとともに、残り 1 ビットをデータ通信の誤り検出（パリティチェックといいます）のために役立てるという工夫が感じられます。

【便利知識】
Windows OS の和文フォントで、ASCII の 16 進数の 5C（本来は半角バックスラッシュ'¥'）の場所に、半角の円記号（'￥'）を割り当ててしまったために、Windows パソコンでは半角バックスラッシュが使いづらいです。
Excel では、その場所だけ和文フォント以外のフォントに変えると半角バックスラッシュで表示できます。
（筆者のパソコンでは、Word、PowerPoint ではこの方法がうまくいきませんでした。全角バックスラッシュ（'＼'）のフォントやフォントサイズを変えるなどで、半角バックスラッシュに見せかけるしかないのかもしれません。ちなみに、'＼'は「すらっしゅ」と入力して変換キーを押せば出てきます。）

ホームページ上では、逆に半角円記号が半角バックスラッシュで見えてしまうことが多いです。特に半角にしなければならない時以外は、円記号もバックスラッシュも、全角にした方が無難です。

日本語用の JIS 規格

米国で生まれたコンピュータは、次いでヨーロッパや日本などに広がり、今では全世界に広がっています。そうすると、各国の言語に対応が必要となり、ASCII では全く不十分になりました。

西欧諸国では、ASCII 文字セットを 8 ビットに拡張し、増えた領域にヨーロッパ諸国で使われる文字を定義した「ISO-8859」という文字セットが使われています。追加する文字の種類によって、西欧諸国で使用する規格、東欧諸国で使用する規格、北欧諸国で使用する規格、ロシア語の規格、など 16 種類の規格が制定されています。

JIS X 0201

日本でも日本工業規格（JIS）によって、同様に ASCII 文字セットを 8 ビットに拡張し、増えた領域に（半角の）カタカナ文字と句読点などの記号を規定した、JIS X 0201 という規格が制定されました。

現在ではこの規格が単独で使用されることはほとんどありませんが、Shift-JIS 文字コード体系や EUC-JP 文字コード体系の一部に組み込まれて用いられています。

16進数	JIS	16進数	JIS	16進数	JIS	16進数	JIS	16進数	JIS	16進数	JIS	16進数	JIS	16進数	JIS
00	NULL	20	SP	40	@	60	`	80		A0	未定義	C0	タ	E0	
01	SOH	21	!	41	A	61	a	81		A1	。	C1	チ	E1	
02	STX	22	"	42	B	62	b	81		A2	「	C2	ツ	E2	
03	ETX	23	#	43	C	63	c	81		A3	」	C3	テ	E3	
04	EOT	24	$	44	D	64	d	81		A4	、	C4	ト	E4	
05	ENG	25	%	45	E	65	e	81		A5	・	C5	ナ	E5	
06	ACK	26	&	46	F	66	f	81		A6	ヲ	C6	ニ	E6	
07	BEL	27	'	47	G	67	g	81		A7	ァ	C7	ヌ	E7	
08	BS	28	(	48	H	68	h	81	未	A8	ィ	C8	ネ	E8	未
09	HT	29	)	49	I	69	i	81		A9	ゥ	C9	ノ	E9	
0A	LF	2A	*	4A	J	6A	j	8A	定	AA	ェ	CA	ハ	EA	定
0B	VT	2B	+	4B	K	6B	k	8B		AB	ォ	CB	ヒ	EB	
0C	FF	2C	,	4C	L	6C	l	8C	義	AC	ャ	CC	フ	EC	義
0D	CR	2D	-	4D	M	6D	m	8D		AD	ュ	CD	ヘ	ED	
0E	SO	2E	.	4E	N	6E	n	8E		AE	ョ	CE	ホ	EE	
0F	SI	2F	/	4F	O	6F	o	8F		AF	ッ	CF	マ	EF	
10	DLE	30	0	50	P	70	p	90		B0	ー	D0	ミ	F0	
11	DC1	31	1	51	Q	71	q	91		B1	ア	D1	ム	F1	
12	DC2	32	2	52	R	72	r	92		B2	イ	D2	メ	F2	
13	DC3	33	3	53	S	73	s	93		B3	ウ	D3	モ	F3	
14	DC4	34	4	54	T	74	t	94		B4	エ	D4	ヤ	F4	
15	NAK	35	5	55	U	75	u	95		B5	オ	D5	ユ	F5	
16	SYN	36	6	56	V	76	v	96		B6	カ	D6	ヨ	F6	
17	ETB	37	7	57	W	77	w	97		B7	キ	D7	ラ	F7	
18	CAN	38	8	58	X	78	x	98		B8	ク	D8	リ	F8	
19	EM	39	9	59	Y	79	y	99		B9	ケ	D9	ル	F9	
1A	SUB	3A	:	5A	Z	7A	z	9A		BA	コ	DA	レ	FA	
1B	ESC	3B	;	5B	[	7B	{	9B		BB	サ	DB	ロ	FB	
1C	FS	3C	<	5C	¥	7C	|	9C		BC	シ	DC	ワ	FC	
1D	GS	3D	=	5D	]	7D	}	9D		BD	ス	DD	ン	FD	
1E	RS	3E	>	5E	^	7E	‾	9E		BE	セ	DE	゛	FE	
1F	US	3F	?	5F	_	7F	DEL	9F		BF	ソ	DF	゜	FF	

※赤は制御文字です（印字されません）
※SP（16進数の20）は半角の空白（ブランク）
※16進数の5Cは、Windows OSの欧文フォントではバックスラッシュ\で表示されます。16進数の7EはASCIIではチルダ（~）ですが、JISではオーバーライン（‾）です。

JIS X 0201 表

JIS X 0208（通称 JIS 漢字コード）

JIS X 0201 では、ひらがなや漢字が扱えません。
漢字のように何千文字もあるような文字体系は、1 バイトのコードでは対応不可能で、2 バイトコード体系の開発が不可欠でした。1 バイトでは最大 256 文字しか入りませんが、2 バイトあれば理論的には最大 65,536 文字まで入ります。2 バイトコードの実用化は日本語や中国語、韓国語の情報処理にとって大きなステップでした。

1978 年、JIS は、2 バイトのコード体系で、ひらがな、漢字（約 6,000 字）、数字や英字の全角文字などを規定する、JIS X 0208 という文字コード体系を規定しました。通称として、JIS 漢字コードとか、JIS 第 1 第 2 水準漢字などと呼ばれるものです。
具体的には、第1バイトと第2バイトの、ともに 16 進数表記で 21～7E のそれぞれ 94 個、全体では 94×94 ＝8,836 個の文字を表すことができる領域に、6,879 文字を割り当てています。

【便利知識】
2000 年に JIS X 0208 コード体系に約 4,000 文字を追加した拡張規格として JIS X 0213 が制定され、2004 年に一部改訂されました。 2000 年版を JIS2000、2004 年版を JIS2004 と称することもあります。

JIS X 0208（JIS 漢字コード）表の一部

シフト JIS（Shift_JIS）

2 バイトコードができたからといって、従来の 1 バイトコードを捨てて、全てが 2 バイトの世界になるわけにはいきません。プログラムのコンパイラやインタプリタなどは、相変わらず ASCII のような 1 バイトコードをベースとして動くので、過去の資産との互換性が求められます。

シフト JIS は、日本の 1 バイトコード体系の JIS X 0201 と 2 バイト漢字コード体系の JIS X 0208 とを組み合わせて運用する方式です。この規格を Shift_JIS といいます。

シフト JIS の漢字コード体系は、JIS 漢字コード体系の表を、8000 代（具体的には 8140 以降）、9000 代（同 9040 以降）、E000 代（同 E040 以降）、F000 代（同 F040 以降）に分散して、移動（シフト）した形になっています。
なぜ、このような位置に漢字コード体系をシフトさせているかというと、JIS X 0201 コード体系で、80～9F、E0～FF の欄が未使用になっているからです。

例えば、「JIS 漢字コード」という文字を、JIS X 0201 表と JIS 漢字コード表を使って、16 進数に変換すると、「4A 49 53 34 41 3B 7A 25 33 21 3C 25 49」となります。
（分かりやすくするため、間に空白を入れました。）
これを、1 バイトコードと思って、JIS X 0201 表で文字に変換していくと、「JIS4A;z%3!<%I」となってしまいます。
同じく、「JIS 漢字コード」という文字を、シフト JIS 漢字コード表を使って、16 進数に変換すると、「4A 49 53 8A BF 8E 9A 83 52 81 5B 83 68」となります。
これを、1バイトコードと思って、JIS X 0201 表で文字に変換しようとすると、「JIS」に変換した後の 16 進数「8A」が未定義であることに気づき、これは 2 バイトコードと解釈して、シフト JIS 漢字コード表を探して「漢」に変換するのです。以下同様に、1 バイトコードなのか、2 バイトコードなのかを判断して、JIS X 0201 表とシフト JIS 漢字コード表を使い分けることで、正しく変換ができるのです。

【便利知識】
JIS X 0208 表の代わりに、文字数が拡張された JIS X 0213 表に準じたシフト JIS コード体系が使用されることもあります。この規格を Shift_JIS-2004 といいます。

【便利知識】
UNIX OS（Web サーバなどの OS として利用されることが多かった）のために開発された日本語用文字コード体系に、EUC-JP という規格があります。これは、ASCII コード体系と JIS X 0208 コード体系を組み合わせたものと言えます。

	0 1 2 3 4 5 6 7 8 9 A B C D E F
8140	、 。 , . ・ : ; ? ! ゛ ゜ ´ ｀ ¨ ^
8150	￣ _ ヽ ヾ ゝ ゞ 〃 仝 々 〆 〇 ー ― ‐ ／ ＼
8160	～ ∥ ｜ … ‥ ‘ ’ “ ” （ ） 〔 〕 ［ ］ ｛
8170	｝ 〈 〉 《 》 「 」 『 』 【 】 ＋ － ± ×
8180	÷ ＝ ≠ ＜ ＞ ≦ ≧ ∞ ∴ ♂ ♀ ° ′ ″ ℃ ￥
8190	＄ ¢ £ ％ ＃ ＆ ＊ ＠ § ☆ ★ ○ ● ◎ ◇ ◆
81A0	□ ■ △ ▲ ▽ ▼ ※ 〒 → ← ↑ ↓ 〓
81B0	∈ ∋ ⊆ ⊇ ⊂ ⊃ ∪ ∩
81C0	∧ ∨ ￢ ⇒ ⇔ ∀ ∃
81D0	∠ ⊥ ⌒ ∂ ∇ ≡
81E0	≒ ≪ ≫ √ ∽ ∝ ∵ ∫ ∬
81F0	Å ‰ ♯ ♭ ♪ † ‡ ¶ ◯
8240	0
8250	１ ２ ３ ４ ５ ６ ７ ８ ９
8260	Ａ Ｂ Ｃ Ｄ Ｅ Ｆ Ｇ Ｈ Ｉ Ｊ Ｋ Ｌ Ｍ Ｎ Ｏ Ｐ
8270	Ｑ Ｒ Ｓ Ｔ Ｕ Ｖ Ｗ Ｘ Ｙ Ｚ
8280	ａ ｂ ｃ ｄ ｅ ｆ ｇ ｈ ｉ ｊ ｋ ｌ ｍ ｎ ｏ
8290	ｐ ｑ ｒ ｓ ｔ ｕ ｖ ｗ ｘ ｙ ｚ ぁ
82A0	あ ぃ い ぅ う ぇ え ぉ お か が き ぎ く ぐ け

	0 1 2 3 4 5 6 7 8 9 A B C D E F
82B0	げ こ ご さ ざ し じ す ず せ ぜ そ ぞ た だ ち
82C0	ぢ っ つ づ て で と ど な に ぬ ね の は ば ぱ
82D0	ひ び ぴ ふ ぶ ぷ へ べ ぺ ほ ぼ ぽ ま み む め
82E0	も ゃ や ゅ ゆ ょ よ ら り る れ ろ ゎ わ ゐ ゑ
82F0	を ん
8340	ァ ア ィ イ ゥ ウ ェ エ ォ オ カ ガ キ ギ ク グ
8350	ケ ゲ コ ゴ サ ザ シ ジ ス ズ セ ゼ ソ ゾ タ ダ
8360	チ ヂ ッ ツ ヅ テ デ ト ド ナ ニ ヌ ネ ノ ハ バ
8370	パ ヒ ビ ピ フ ブ プ ヘ ベ ペ ホ ボ ポ マ ミ
8380	ム メ モ ャ ヤ ュ ユ ョ ヨ ラ リ ル レ ロ ヮ ワ
8390	ヰ ヱ ヲ ン ヴ ヵ ヶ Α
83A0	Β Γ Δ Ε Ζ Η Θ Ι Κ Λ Μ Ν Ξ Ο Π Ρ
83B0	Σ Τ Υ Φ Χ Ψ Ω α
83C0	β γ δ ε ζ η θ ι κ λ μ ν ξ ο π ρ
83D0	σ τ υ φ χ ψ ω
83E0	
83F0	
8440	А Б В Г Д Е Ё Ж З И Й К Л М Н О
8450	П Р С Т У Ф Х Ц Ч Ш Щ Ъ Ы Ь Э Ю
8460	Я

中略

	0 1 2 3 4 5 6 7 8 9 A B C D E F
88A0	唖 娃 阿 哀 愛 挨 姶 逢 葵 茜 穐 悪 握 渥 旭 葦
88B0	芦 鯵 梓 圧 斡 扱 宛 姐 虻 飴 絢 綾 鮎 或 粟 袷
88C0	安 庵 按 暗 案 闇 鞍 杏 以 伊 位 依 偉 囲 夷 委
88D0	威 尉 惟 意 慰 易 椅 為 畏 異 移 維 緯 胃 萎 衣
88E0	謂 違 遺 医 井 亥 域 育 郁 磯 一 壱 溢 逸 稲 茨
88F0	芋 鰯 允 印 咽 員 因 姻 引 飲 淫 胤 蔭
8940	院 陰 隠 韻 吋 右 宇 烏 羽 迂 雨 卯 鵜 窺 丑 碓
8950	臼 渦 嘘 唄 鬱 蔚 鰻 姥 厩 浦 瓜 閏 噂 云 運 雲
8960	荏 餌 叡 営 嬰 影 映 曳 栄 永 泳 洩 瑛 盈 穎 頴
8970	英 衛 詠 鋭 液 疫 益 駅 悦 謁 越 閲 榎 厭 円
8980	園 堰 奄 宴 延 怨 掩 援 沿 演 炎 焔 煙 蒸 猿 縁
8990	艶 苑 薗 遠 鉛 鴛 塩 於 汚 甥 凹 央 奥 往 応 押
89A0	旺 横 欧 殴 王 翁 襖 鴬 鴎 黄 岡 沖 荻 億 屋 憶
89B0	臆 桶 牡 乙 俺 卸 恩 温 穏 音 下 化 仮 何 伽 価
89C0	佳 加 可 嘉 夏 嫁 家 寡 科 暇 果 架 歌 河 火 珂
89D0	禍 禾 稼 箇 花 苛 茄 荷 華 菓 蝦 課 嘩 貨 迦 過
89E0	霞 蚊 俄 峨 我 牙 画 臥 芽 蛾 賀 雅 餓 駕 介 会
89F0	解 回 塊 壊 廻 快 怪 悔 恢 懐 戒 拐 改
8A40	魁 晦 械 海 灰 界 皆 絵 芥 蟹 開 階 貝 凱 劾 外
8A50	咳 害 崖 慨 概 涯 碍 蓋 街 該 鎧 骸 浬 馨 蛙 垣
8A60	柿 蛎 鈎 劃 嚇 各 廓 拡 撹 格 核 殻 獲 確 穫 覚

	0 1 2 3 4 5 6 7 8 9 A B C D E F
8A70	角 赫 較 郭 閣 隔 革 学 岳 楽 額 顎 掛 笠 樫
8A80	橿 梶 鰍 潟 割 喝 恰 括 活 渇 滑 葛 褐 轄 且 鰹
8A90	叶 椛 樺 鞄 株 兜 竃 蒲 釜 鎌 噛 鴨 栢 茅 萱 粥
8AA0	刈 苅 瓦 乾 侃 冠 寒 刊 勘 勧 巻 喚 堪 姦 完 官
8AB0	寛 干 幹 患 感 慣 憾 換 敢 柑 桓 棺 款 歓 汗 漢
8AC0	澗 潅 環 甘 監 看 竿 管 簡 緩 缶 翰 肝 艦 莞 観
8AD0	諌 貫 還 鑑 間 閑 関 陥 韓 館 舘 丸 含 岸 巌 玩
8AE0	癌 眼 岩 翫 贋 雁 頑 顔 願 企 伎 危 喜 器 基 奇
8AF0	嬉 寄 岐 希 幾 忌 揮 机 旗 既 期 棋 棄
8B40	機 帰 毅 気 汽 畿 祈 季 稀 紀 徽 規 記 貴 起 軌
8B50	輝 飢 騎 鬼 亀 偽 儀 妓 宜 戯 技 擬 欺 犠 疑 祇
8B60	義 蟻 誼 議 掬 菊 鞠 吉 吃 喫 桔 橘 詰 砧 杵 黍
8B70	却 客 脚 虐 逆 丘 久 仇 休 及 吸 宮 弓 急 救
8B80	朽 求 汲 泣 灸 球 究 窮 笈 級 糾 給 旧 牛 去 居
8B90	巨 拒 拠 挙 渠 虚 許 距 鋸 漁 禦 魚 亨 享 京 供
8BA0	侠 僑 兇 競 共 凶 協 匡 郷 叫 喬 境 峡 強 彊 怯
8BB0	恐 恭 挟 教 橋 況 狂 狭 矯 胸 脅 興 蕎 郷 鏡 響
8BC0	饗 驚 仰 凝 尭 暁 業 局 曲 極 玉 桐 粁 僅 勤 均
8BD0	巾 錦 斤 欣 欽 琴 禁 禽 筋 緊 芹 菌 衿 襟 謹 近
8BE0	金 吟 銀 九 倶 句 区 狗 玖 矩 苦 躯 駆 駈 駒 具
8BF0	愚 虞 喰 空 偶 寓 遇 隅 串 櫛 釧 屑 屈

後略

シフト JIS 漢字コード表（一部）

【便利知識】
シフト JIS は、マイクロソフト社と日本企業数社が協力して 1982 年に誕生しました。そして、MS 漢字コードとも呼ばれ、Windows OS（日本語版）の標準として活躍してきましたが、一方で、インターネットで世界中が繋がり、Web サイトなどを通じて多数の言語に触れる機会が増し、Web サイトなどの閲覧、メールや Word 等の文書データの閲覧時に、しばしば文字化けが発生するようになってきました。

シフト JIS 自体も、もともと未使用の 1 バイトコードがあったことを利用したアクロバティカルな方式で、登録されている文字数に限りがあることと、文字が規定されていない空きエリアを利用して、種々の外字が登録されていき、多数の亜流の文字コード体系が生まれてしまい、標準としては限界に近づいています。

このような背景と、後述のように、世界共通の標準的なコード体系の制定という動きに応じて、まもなく、シフト JIS は Windows の標準ではなくなる運命にあります。

UTF-8

世界で使われる全ての文字を共通の文字コード体系で利用できるようにしようという考えから、ゼロックス、マイクロソフト、アップル、IBM、サン・マイクロシステムズ、ヒューレット・パッカード、ジャストシステムなどが参加したコンソーシアムによって、Unicode（ユニコード）という符号化文字集合や文字符号化方式などを定めた文字コードの業界規格が、1991 年に誕生しました。

単一の大規模な文字セットであることが特徴で、世界各国の現代の文字だけでなく、古代の文字や歴史的な文字、数学記号、絵文字（ガラケーで使用されているものも）などにも対応しています。
国際規格の ISO/IEC 10646 と Unicode 規格は同じ文字コード表になるように協調して策定されています。

Unicode は、Unix、Windows、macOS、Java など種々の OS で利用されており、文字符号化方式も標準化されていて、いくつかの亜種があった Shift_JIS や EUC-JP の混乱のようなものは回避されています。

Unicode は当初、1 文字を 2 バイト（16 ビット）の固定長で表す規格のコード化方式で考えられ、UTF-16（Unicode Transformation Format-16）と名づけられました。ただ 2 バイトでは、最大で約 6 万字の文字しかサポートできないので、結局、2 バイト固定長という原則を崩すことになりました。基本的には（名前の通り）2 バイトで 1 文字を表

すのですが、「サロゲート」と呼ばれる拡張領域内の文字を利用する場合は 1 文字を 4 バイトで表します。つまり、UTF-16 は 2 バイト表現と 4 バイト表現が混在したコード変換方式です。

UTF-16 は ASCII 文字も 1 文字 2 バイトで表現します。つまり ASCII コードと互換性がありません。そのため、ASCIIコードを想定したプログラムで誤動作を起こす可能性があります。そこで ASCII 文字については、ASCII と同様のコードを割り当てるようにしたコード変換方式が考えられました。それが UTF-8 です。

UTF-8 は 1～4 バイトで 1 文字を表し、漢字や仮名の 1 文字は 3 バイトを要します。シフト JIS では漢字や仮名は2バイトでしたので、日本語の文書などでは、データサイズが 1.5 倍程度に増大することになります。
Unicode には他に 4 バイト固定長で処理する UTF-32 や UCS-4 などの方式も規定されています。

【便利知識】
上述までの解説でお分かりのように、現在の文字コード変換の標準は、UTF-8 に集約されつつあります。文字コードは UTF-8 にするという考え方でいけば、文字化けも起こりにくいと言えましょう。