

情報システム論

第4回

今日の話題

- 文字コード

集合

{ 晴れ, 雨, 曇り } = { 晴れ, 曇り, 雨 }

{ 晴れ, 雨, 曇り, 雪 } = { 晴れ, 雪, 曇り, 雨 }

{ ☀, ☂, ☁, ☷ } = { ☁, ☀, ☂, ☷ }

符号化集合

{ 晴れ, 雨, 曇り }

1 2 3

{ 晴れ, 雨, 曇り }

1 2 3

{ 晴れ, 雨, 曇り, 雪 }

1 2 3 4

{ 晴れ, 雪, 曇り, 雨 }

1 2 3 4

{ , , ,  }

1 2 3 4

{ , , ,  }

1 2 3 4

符号化集合

{ ☀, ☂, ☁, ☺ }

⇒ { 1, 2, 3, 4 }

⇒ { (00)₂, (01)₂, (10)₂, (11)₂ }

{ !, ", #, \$, ..., 0, 1, 2, 3, ..., A, B, C, ..., a, b, c, ... }

⇒ { 1, 2, 3, 4, ..., 11, 12, 13, 14, ..., 31, 32, 33, ... 61, 62, 63, ... }

⇒ { (21)₁₆, (22)₁₆, (23)₁₆, (24)₁₆, ..., (30)₁₆, (31)₁₆, (32)₁₆, (33)₁₆, ...,
(41)₁₆, (42)₁₆, (43)₁₆, ..., (61)₁₆, (62)₁₆, (63)₁₆, ... }

{ ●, ◎, ◇, ..., あ, あ, い, い, ..., 亜, 啞, 娃, ..., 湾, 碗, 腕, ... }

⇒ { 1, 2, 3, ..., 94, 95, 96, 97, ..., 567, 568, 569, ..., 4132, 4133, 4134, ... }

⇒ { (217C)₁₆, (217D)₁₆, (217E)₁₆, ..., (2421)₁₆, (2422)₁₆, (2423)₁₆, (2422)₁₆, ...,
(3021)₁₆, (3022)₁₆, (3023)₁₆, ..., (4F51)₁₆, (4F52)₁₆, (4F53)₁₆ }

符号化方式

{ 晴れ, 雨, 曇り }

1	2	3
01	10	11

{ 晴れ, 雨, 曇り, 雪, 霧, 雹, 霰 }

1	2	3	4	5	6	7
000	001	010	011	100	101	110
00	01	10	1100	1101	1110	111101

符号化文字集合

- ISO/IEC 646
 - ASCII
 - JIS X 0201
- JIS X 0208
- Unicode

ASCII

- 1963年にAmerican Standards Association (ASA、後のANSI)で制定
- 7ビットコード
- コード範囲
 - 00-1F 制御文字
 - 20 空白
 - 21-7E 図形文字
 - 7F 制御文字

	0-	1-	2-	3-	4-	5-	6-	7-
0			SP	0	@	P	`	p
-1			!	1	A	Q	a	q
-2			“	2	B	R	b	r
-3			#	3-	C	S	c	s
-4			\$	4	D	T	d	t
-5			%	5	E	U	e	u
-6			&	6	F	V	f	v
-7			‘	7	G	W	g	w
-8			(8	H	X	h	x
-9)	9	I	Y	i	y
-A			*	:	J	Z	j	z
-B			+	;	K	[k	{
-C			,	<	L	\	l	
-D			-	=	M]	m	}
-E			.	>	N	^	n	~
-F			/	?	O	_	o	

ISO/IEC 646

- ASCIIを元に行している
- 7ビットコード
- コード範囲
 - 00-1F 制御文字
 - 20 空白
 - 21-7E 図形文字
 - 7F 制御文字

	0-	1-	2-	3-	4-	5-	6-	7-
0			SP	0	@	P	`	p
-1			!	1	A	Q	a	q
-2			“	2	B	R	b	r
-3			#	3-	C	S	c	s
-4			\$	4	D	T	d	t
-5			%	5	E	U	e	u
-6			&	6	F	V	f	v
-7			‘	7	G	W	g	w
-8			(8	H	X	h	x
-9)	9	I	Y	i	y
-A			*	:	J	Z	j	z
-B			+	;	K	[k	{
-C			,	<	L	\	l	
-D			-	=	M]	m	}
-E			.	>	N	^	n	~
-F			/	?	O	_	o	

JIS X 0201 (7ビット及び8ビットの 情報交換用符号化文字集合)

	0-	1-	2-	3-	4-	5-	6-	7-	8-	9-	A-	B-	C-	D-	E-	F-
0			SP	0	@	P	`	p				ー	タ	ミ		
-1			!	1	A	Q	a	q			。	ア	チ	ム		
-2			“	2	B	R	b	r			「	イ	ツ	メ		
-3			#	3-	C	S	c	s			」	ウ	テ	モ		
-4			\$	4	D	T	d	t			、	エ	ト	ヤ		
-5			%	5	E	U	e	u			・	オ	ナ	ユ		
-6			&	6	F	V	f	v			ヲ	カ	ニ	ヨ		
-7			‘	7	G	W	g	w			ア	キ	ヌ	ラ		
-8			(8	H	X	h	x			イ	ク	ネ	リ		
-9)	9	I	Y	i	y			ウ	ケ	ノ	ル		
-A			*	:	J	Z	j	z			エ	コ	ハ	レ		
-B			+	;	K	[k	{			オ	サ	ヒ	ロ		
-C			,	<	L	¥	l				ヤ	シ	フ	ワ		
-D			-	=	M]	m	}			ユ	ス	ヘ	ン		
-E			.	>	N	^	n	—			ヨ	セ	ホ	”		
-F			/	?	O	_	o				ツ	ソ	マ	°		

JIS X 0201 (7ビット及び8ビットの 情報交換用符号化文字集合)

	0-	1-	2-	3-	4-	5-	6-	7-	8-	9-	A-	B-	C-	D-	E-	F-
0			SP	0	@	P	`	p				ー	タ	ミ		
-1			!	1	A	Q	a	q			。	ア	チ	ム		
-2			“	2	B	R	b	r			「	イ	ツ	メ		
-3			#	3-	C	S	c	s			」	ウ	テ	モ		
-4			\$	4	D	T	d	t			、	エ	ト	ヤ		
-5			%	5	E	U	e	u			・	オ	ナ	ユ		
-6			&	6	F	V	f	v			ヲ	カ	ニ	ヨ		
-7			‘	7	G	W	g	w			ア	キ	ヌ	ラ		
-8			(8	H	X	h	x			イ	ク	ネ	リ		
-9)	9	I	Y	i	y			ウ	ケ	ノ	ル		
-A			*	:	J	Z	j	z			エ	コ	ハ	レ		
-B			+	;	K	[k	{			オ	サ	ヒ	ロ		
-C			,	<	L	¥	l				ヤ	シ	フ	ワ		
-D			-	=	M]	m	}			ユ	ス	ヘ	ン		
-E			.	>	N	^	n	—			ヨ	セ	ホ	”		
-F			/	?	O	_	o				ツ	ソ	マ	°		

JIS X 0201 (7ビット及び8ビットの 情報交換用符号化文字集合)

	0-	1-	2-	3-	4-	5-	6-	7-
0			SP	0	@	P	`	p
-1			!	1	A	Q	a	q
-2			“	2	B	R	b	r
-3			#	3-	C	S	c	s
-4			\$	4	D	T	d	t
-5			%	5	E	U	e	u
-6			&	6	F	V	f	v
-7			‘	7	G	W	g	w
-8			(8	H	X	h	x
-9)	9	I	Y	i	y
-A			*	:	J	Z	j	z
-B			+	;	K	[k	{
-C			,	<	L	¥	l	
-D			-	=	M]	m	}
-E			.	>	N	^	n	—
-F			/	?	O	_	o	

	0-	1-	2-	3-	4-	5-	6-	7-
0				一	夕	ミ		
-1			。	ア	チ	ム		
-2			「	イ	ツ	メ		
-3			」	ウ	テ	モ		
-4			、	エ	ト	ヤ		
-5			・	オ	ナ	ユ		
-6			ヲ	カ	ニ	ヨ		
-7			ア	キ	ヌ	ラ		
-8			イ	ク	ネ	リ		
-9			ウ	ケ	ノ	ル		
-A			エ	コ	ハ	レ		
-B			オ	サ	ヒ	ロ		
-C			ヤ	シ	フ	ワ		
-D			ユ	ス	ヘ	ン		
-E			ヨ	セ	ホ	”		
-F			ツ	ソ	マ	°		

Unicode

- 世界中の文字の16進で4桁以上の数字を割当てたもの
- U+XXXX で表現
 - U+3001 、
 - U+3001 。
 - U+3041 あ
 - U+3042 あ

符号化方式

- 符号化集合の文字に対応づけられた整数に対して、コンピュータが利用できるよデータ列に変換する方式
 - JIS X 0201
 - JIS X 0208
 - ISO-2022-JP
 - EUC-JP
 - Shift_JIS
 - Unicode
 - UTF-16
 - UTF-8

漢字用7ビット符号

- JIS X 0208 の区点(1-94,1-94)を(21)_H-(7E)_Hに配置

- 愛(区点1606)

30	26
----	----

- 愛0&

30	26	30	26
----	----	----	----

EUC-JP

- JIS X 0208 の区点(1-94,1-94)を(A1)_H-(EE)_Hに配置

- 愛(区点1606)

B0	A6
----	----

- 愛0&

B0	A6	30	26
----	----	----	----

Shift JIS

- JIS X 0208 の
 - 区(1-94)を(81)_H-(9F)_H と (E0)_H-(EF)_H
 - 点(1-94)を(40)_H-(7E)_Hと (80)_H-(FC)_H

- 愛0&

88	A4	30	26
----	----	----	----

ISO 2022-JP

- エスケープシーケンス

- ASCII IB 28 42

- JIS X 0208-1983 IB 24 42

- 愛0&

IB	24	42	30	26	IB	28	42	30	26
----	----	----	----	----	----	----	----	----	----

文字化け

- Encodingの誤認識
- 未定義エリアへの独自文字の割当
 - 表示されない
 - 別の文字が表示

包摂・字形変更

- 包摂
 - 複数の字体を区別せずに同一のコード
- 字形変更
 - 鷗(JIS C 6226-1978)⇒鷗(JIS C 6226-1983)
 - 壺(第一水準) ⇔ 壺(第二水準)

改行コード

- Windows
 - CR+LF
- Mac
 - CR
- Unix
 - LF

全角・半角

- もともと印刷用語
 - 2byte文字が全角、1byte文字が半角を占めていた
- A A
- 検索時の問題

リンク

- <http://ja.wikipedia.org/wiki/ASCII>
- http://ja.wikipedia.org/wiki/ISO/IEC_646
- http://ja.wikipedia.org/wiki/JIS_X_0201
- http://ja.wikipedia.org/wiki/JIS_X_0208
- <http://ja.wikipedia.org/wiki/Unicode>
- <http://ja.wikipedia.org/wiki/ISO-2022>