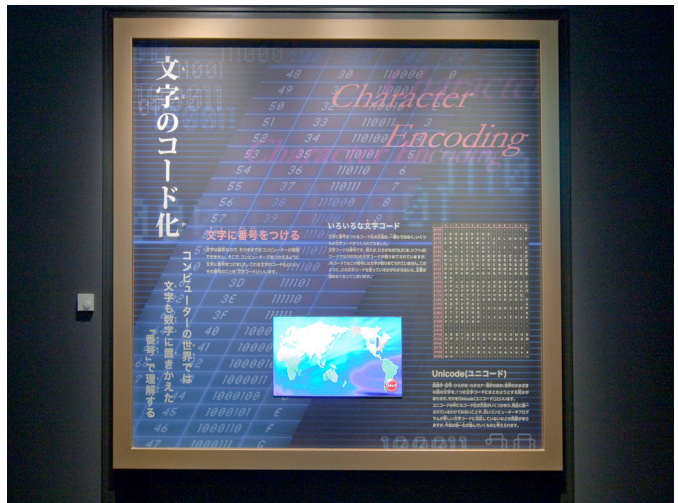# S430 Character Encoding

文字のコード化

■Purpose of Exhibition

Inside a computer, all the characters are numbered. Numbering characters is called "Encoding characters" and some numbered characters are referred to as "Character codes". In this exhibition, try to decode numbers to characters in the game.

■Additional Knowledge

	first digit of a number					
	20	30	40	50	60	70
0		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	

Chart.1　ASCII code (abridgment)

_US ASCII codeIn the U.S., where the computer was first produced, 128 characters were encoded in the first place. These are called US ASCII codes (Chart1). In addition to upper and lower cases in alphabet, numbers, and symbols, there are some character codes with functions that cannot be shown on display such as the delete signs on a display or a sound like "pip" emitting from speakers.

_65535Since there are some tens of thousands of kinds of characters including hiragana, katakana, and kanji in Japanese, character codes, which enable 65535 kinds of characters to show, are in use.

But why does it have to be 65535? Not simply like 10 thousand or 100 thousand? That is due to the way codes are shown. Although calculation is done with a binary number system inside a computer, it becomes to long to show bigger numbers with the binary number system. In order for it to become more convenient, hexadecimal is used. To show 65535 kinds of characters, numbers from 0 to 65535 are utilized. There are 65535 from 0 to 65535."0" is included also, therefore there are the numbers 0 to 65535. In hexadecimal, this can be indicated as "from 0x0 to 0xFFFF". The range that can show two-digit numbers in decimal is 0 to 65535. It means to indicate numbers with 2 bytes. These characters are called 2 bytes characters.

_Various CodesMany kinds of codes lie in character codes and this causes inconvenience at times. The character code adopted in Microsoft products is called "Shift JIS". This is something that has made a small change in "JIS codes". "EUC" was common in computer basic software called "UNIX". Other than that, there are ones like "Unicode", "UTF", which deal with Japanese, "EBCDIC", which was previously well-used, and more. In short, it indicates that numbers can change depending on with what codes numbers can be presented despite the same code. Since a computer deals only with numbers, a phenomenon occurs: if a word "a" (in Japanese) is dealt with by other character codes, an entirely different character can be obtained. This is called "Garbled character".

_E-mailsRecently, e-mails have been in use on personal computers and in mobile phones. Although the structure between a mobile phone and a personal computer should be disparate, e-mails are sent without being garbled in most cases. How does it works? The answer lies in a structure of e-mails. An e-mail has a part called a "header" outside of the main text. There is a designation that shows which character encoding a text should be read with. Besides that, a header has clauses such as a sender, date, recipient, subject, mailing record, and mailing software in use. E-mails come in handy with secrets like these to transmit information without any failure.

Article by Tetsuo Ojio, curator