Approval: 10/21/65

Identification

Character set
F.J. Corbato, R. Morris, 7/23/65

Purpose

The system will use a standard character set which will be the newly revised ASCII character set (Reference: Communications of the ACM, April 1965, p. 207). The code is a 7-bit code (mapped into a 9-bit field in the GE636) containing 128 characters. Any devices which cannot create or accept the full character set will use established conventions for representing the full set. It is emphasized that there are no meaningful subsets of the revised ASCII character set.

Revised ASCII character set

The ASCII character set consists of 128 characters, 95 of which have associated graphics, and 33 of which are control characters. The graphics include A - Z, a - z, 0 - 9, space
! " # $ % & / ( ) * + , - . ~ : ; ' ` = ? <> { } ¬ | _ @ ^ [ ]

The Control Characters are:

NUL	Null	DC3	Device Control 3
SOH	Start of Heading (CC)	DC4	Device Control (stop)
STX	Start of Text (CC)	NAK	Negative Acknowledge (CC)
ETX	End of Text (CC)	SYN	Synchronous Idle (CC)
EOT	End of Transmission (CC)	ETB	End of Transmission Block (CC)
ENQ	Enquiry (CC)	CAN	Cancel
ACK	Acknowledge (CC)	EM	End of Medium
BEL	Bell (audible or attention signal)	SS	Start of Special Sequence
BS	Backspace (FE)	ESC	Escape
HT	Horizontal Tabulation (punched card skip) (FE)	FS	File Separator (IS)
LF	Line Feed (FE)	GS	Group Separator (IS)
VT	Vertical Tabulation (FE)	RS	Record Separator (IS)
FF	Form Feed (FE)	US	Unit Separator (IS)
CR	Carriage Return (FE)	DEL	Delete*
SO	Shift Out		
SI	Shift In		
DLE	Data Link Escape (CC)		
DC1	Device Control 1		
DC2	Device Control 2		

(CC) Communication Control. (FE) Format Effector.
(IS) Information Separator.
* In the strict sense, DEL is not a control character.

Standard 7-bit Character Set

b765	000	001	010	011	100	101	110	111
b4321								
0000	NUL	DLE	SP	0	\	P	@	p
0001	SOH	DC1	!	1	A	Q	a	q
0010	STX	DC2	"	2	B	R	b	r
0011	ETX	DC3	#	3	C	S	c	s
0100	EOT	DC4	$	4	D	T	d	t
0101	ENQ	NAK	%	5	E	U	e	u
0110	ACK	SYN	&	6	F	V	f	v
0111	BEL	ETB	'	7	G	W	g	w
1000	BS	CAN	(	8	H	X	h	x
1001	HT	EM	)	9	I	Y	i	y
1010	LF	SS	*	:	J	Z	j	z
1011	VT	ESC	+	;	K	[	k	{
1100	FF	FS	,	<	L	~	l	|
1101	CR	GS	-	=	M	]	m	}
1110	SO	RS	.	>	N	^	n	~
1111	SI	US	/	?	O	_	o	DEL

Example:

The bit representation for the character "K" is:

b7	b6	b5	b4	b3	b2	b1
1	0	0	1	0	1	1

Standard 7-bit ASCII Character Set (alternate table format)

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

Standard Escape Conventions

Contemporary console equipment often is not capable of representing all 128 of the code values. To keep full generality and flexibility in the future, standard software escape conventions will be used for all console devices. On devices which have the revised ASCII set, the use of the escape mechanism will normally be unnecessary. Each class of console device has a particular character assigned as the "software escape" character; This character when it occurs in an input (or output) string to (or from) a console, always gives a special interpretation to the next one or more characters. The standard escape character will be the circumflex this means that to input the code for it, an escape convention will have to be used. Therefore the circumflex should be avoided in all multics software. (It should be noted that the two standard "kill characters" @ and # should also be avoided in all software.)

All Consoles

For simplicity, escape conventions have been established which are uniform over several possible console classes. For full flexibility there is a mechanism for representing any arbitrary octal code in a character string. This is:

> ^d1d2d3 for the octal code d1 d2 d3 (modulo 128 on input modulo 512 on output) where d1, d2, d3 are from 0 to 7.
>
> ^Ck for a local (i.e. concealed) use of the character k which does not go into the computer-stored string on input and which is not in the computer-stored string on output. (For example, a concealed carriage return given to avoid a right-margin jam up.)

Standard left margin and horizontal tab settings, which should be assumed by all multics software to avoid the nuisance of special tab settings, are at the following locations: 0,10,...,130. (Software conventions may be established later which allow declarations by the user to the system of tab setting changes thereby allowing arbitrary tabs for special purposes.)

37KSR Teletypes

There are no further escape conventions required of the model 37KSR Teletype since it has the revised ASCII character set.

IBM 1050 Console

Each type ball used will require a different set of escape conventions. The standard "correspondence" type ball is currently the only one available.

Correspondence ball:

The following non-revised ASCII characters on the ball are considered to be stylized versions of ASCII characters:

¤	(lozenge)	for	^ (circumflex - software excape)
¢	(cent sign)	for	| (vertical stroke)
±	(plus-minus)	for	~ (tilde)
´	(apostrophe)	for	' (accent acute)

In addition to the two universal escape conventions, the following are available for convenience:

¤'	for	` (accent grave)
¤n	for	¬ (logical negation)
¤¤(	for	[(left square bracket)
¤¤)	for	] (right square bracket)
¤l	for	< (less than)
¤g	for	> (greater than)
¤(	for	{ (left brace)
¤)	for	} (right brace)

Project MAC Type Ball: (not available until after Jan. 1966)

The following non-revised ASCII characters on the ball are considered to be stylized versions of ASCII characters:

\	(reverse slash)	for	^ (circumflex - software escape)
'	(apostrophe)	for	' (accent acute)

In addition to the two universal escape conventions, the following are available for convenience:

\ '	for	`	(accent grave)
\ p	for	%	(percent)
\ .	for	!	(exclamation)
\ (	for	{	(left brace)
\ )	for	}	(right brace)

35KSR and 33KSR Teletypes and 64 Character "ASCII" Keypunches

The escape character is: \ (reverse slash). The following non-revised ASCII characters are considered to be stylized versions of ASCII characters.

\	(reverse slash)	for	^ (circumflex - software escape)
↑	(vertical arrow)	for	| (vertical stroke)

← (left arrow)	for	—	(underline)
' (apostrophe)	for	´	(accent acute)

In addition to the two universal escape conventions, the following are available for convenience:

\ - designates backspace.

\ U designates that all subsequent letters are in upper case until a \L sequence is encountered. (This is assumed as the initial letter case mode.)

\ L designates that all subsequent letters are in lower case until a \U sequence is encountered.

\\ n where n = 1, 2, ..., 9 designates that the next alphabetic characters are to be in the opposite case from the present case. (This is useful for initial capitilization of words, etc.)

\ N	for	¬	(logical negation)
\ T	for	~	(tilde)
\ '	for	`	(accent grave)
\ (	for	{	(left brace)
\ )	for	}	(right brace)

In the case of key punches, an end-of-card automatically generates a carriage return it is also convenient for input to have:

\ * for "skip reading the remainder of this card without a carriage return"

\ / for "carriage return and skip reading the remainder of this card"

\ + for "carriage return and keep reading this card"

\ H for horizontal tab