

TO: MSPM Distribution
FROM: J. H. Saltzer
SUBJECT: BC.2.01
DATE: 09/25/68

The enclosed section BC.2.01 incorporates a data padding character and takes into account anticipated wide use of the ASCII Null character outside Multics. Appendix BC.2.01A, 09/03/68, is also superseded by this version. That appendix was incorrect and anyone making use of it is advised to review the corresponding parts of this section carefully.

Published: 09/25/68
 (Supersedes: BC.2.01, 02/09/68,
 BC.2.01, 02/05/68,
 BC.2.01, 11/08/66,
 BC.2.01, 08/17/66,
 BC.2.01, 01/27/66,
 BC.2, 09/24/65,
 BC.2, 07/23/65)

Identification

Character Set
 F. J. Corbat⁶, R. Morris, J. H. Saltzer

Purpose

The system shall use a standard character set which is the revised ASCII character set (Reference: USA Standards Institute, "USA Standard X3.4-1967"). The ASCII set consists of 128 7-bit characters which shall be right adjusted with two leading zeros in the GE 645. The two leading bit positions are expressly reserved for future character set expansion and shall not be used by any system program. Any devices which cannot create or accept the full character set shall 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

Included in the ASCII character set are 94 printing graphics, space, and 33 control characters. Within Multics, all the graphics, space, and 11 of the control characters are given precise interpretations. The remaining 22 control characters are presently reserved. The graphics in the set are the following:

Upper Case Alphabet
 ABCDEFGHIJKLMNOPQRSTUVWXYZ

Lower Case Alphabet
 abcdefghijklmnopqrstuvwxyz

Numerals
 1234567890

Special characters
 ! Exclamation Point
 " Double Quote Mark
 # Number Sign
 \$ Dollar Sign
 % Percent
 & Ampersand
 ' Acute Accent
 (Open Parenthesis
) Close Parenthesis

; Semicolon
 < Less Than
 = Equals
 > Greater Than
 ? Question Mark
 @ Commercial At
 [Open Bracket
 \ Left Slant
] Close Bracket

* Asterisk	^ Circumflex
+ Plus	_ Underline
, Comma	` Grave Accent
- Minus	{ Open Brace
. Period	Vertical Line
/ Slant	} Close Brace
: Colon	~ Overline/Tilde

Control Characters.

The following conventions define the only standard meaning for the ASCII control characters within Multics. These conventions will be observed by all standard Device Interface Modules (DIM'S) and by all system software inside the I/O system interface. Since some hardware devices will have different interpretations for some characters, the appropriate DIM must perform any necessary translations. Those characters noted as "not used" are reserved; future definitions may be made at any time. Until defined, "not used" control characters encountered on output will be printed out with the octal escape convention described in section BC.2.04. Users wishing a different convention for interpretation of a "not used" character must use a non-standard DIM.

If a device does not allow a function implied by a certain control character, its standard DIM will provide some reasonable interpretation for the character on output, possibly by ignoring the character.

The Multics Standard Control Characters are:

BEL	Sound an audible alarm.
BS	Move carriage back one column. The backspace character implies overstriking rather than erasure.
HT	Horizontal Tabulate. Move carriage to next horizontal tab stop. On variable tab machines, in the absence of an explicit agreement between the typist and his DIM, horizontal tab stops are assumed to be at columns 11, 21, 31, 41, etc. This character cannot appear in a canonical character string.
NL	New Line. Move carriage to left edge of next line. Character code 012 (ASCII LF) is used for this function.
RRS	Red Ribbon Shift. Character code 016 (ASCII S0) is used for this function.

- BRS Black Ribbon Shift.
Character code 017 (ASCII SI) is used for this function.
- VT Vertical Tabulate. Move carriage to next vertical tab stop. On variable-tab machines, in the absence of an explicit agreement between the typist and his DIM, vertical tab stops are assumed to be at lines 11, 21, 31, etc. This character cannot appear in a canonical character string.
- NP New Page. Move carriage to top of the next page. If a graphic follows, it will be printed on the first line of the page, at the left edge.
Character code 014 (ASCII FF) is used for this function.
- HLF Half-Line Forward Feed.
Character code 022 (ASCII DC2) is used for this function.
- HLR Half-Line Reverse Feed.
Character code 024 (ASCII DC4) is used for this function.
- PAD Padding Character. Used to fill out words which contain fewer than four characters and which are not accompanied by character counts. This character is discarded when encountered in an output line. This character cannot appear in a canonical character string.
Character code 177 (ASCII DEL) is used for this function.

Non Standard Control Character

One control character is recognized under certain conditions by all Device Interface Modules, because of its wide use outside of Multics. This character is handled specially only when the device interface module has been conditioned to produce "edited" output, ignoring certain unavailable control functions. (Section BF.1.02 describes the "edited" output mode.) This character is

- NUL Null Character (Code 000) In "edited" output mode, this character is treated exactly as PAD, above. In normal output mode it is considered a "not used" character, and printed with an octal escape sequence. This character cannot appear in a canonical character string.

Programmers are warned against using the NUL character as a routine padding character, because the ensuing use of edited mode output will cause all strings of zeros (including character strings mistakenly never initialized) to be discarded.

Not Used:

SOH(001)	STX(002)	ETX(003)	EOT(004)	ENQ(005)	ACK(006)
CR(015)	DLE(020)	DC1(021)	DC3(023)	NAK(025)	SYN(026)
ETB(027)	CAN(030)	EM(031)	SUB(032)	ESC(033)	FS(034)
GS(035)	RS(036)	US(037)			

On the following page is a table of the ASCII character set as used on Multics. Positions occupied by unused control character codes are blank.

Note:

The vertical line character has two alternative graphic representations on current console devices. It may be represented either as a solid vertical line (|) or as a broken vertical line (|). These two graphic representations are considered to represent the same ASCII character, code value 174.

ASCII Character Set on Multics

	0	1	2	3	4	5	6	7
000	(NUL)							BEL
010	BS	HT	NL	VT	NP		RRS	BRS
020			HLF		HLR			
030								
040	Space	!	"	#	\$	%	&	'
050	()	*	+	,	-	.	/
060	0	1	2	3	4	5	6	7
070	8	9	:	;	<	=	>	?
100	@	A	B	C	D	E	F	G
110	H	I	J	K	L	M	N	O
120	P	Q	R	S	T	U	V	W
130	X	Y	Z	[\]	^	_
140	`	a	b	c	d	e	f	g
150	h	i	j	k	l	m	n	o
160	p	q	r	s	t	u	v	w
170	x	y	z	{		}	~	PAD

Multics Definitions:

NL	New Line (carriage return and line feed)
HLF	Half-Line Forward Feed
HLR	Half-Line Reverse Feed
RRS	Red Ribbon Shift
BRS	Black Ribbon Shift
NP	New Page (carriage return and form feed)
PAD	Padding character