Identification

Typewriter code conversion
D. M. Ritchie

Purpose

This section can be regarded as an addendum to BC.2, which also describes typewriter code conversion. The general principles described therein remain correct, but there are some differences in detail.

Remark

This section tells it like it is.

General

Three types of device are handled: high-speed M37 Teletypes, and IBM 1050 and 2741 typewriters. Low-speed (older) M37 Teletypes will work well enough to be usable. The IBM typewriters should have type 963 golf balls, but a half-hearted attempt is made to accommodate 938 balls.

Two forms of output conversion are provided: edited and normal.

I/O system order calls are provided to change golf balls on a 1050 or 2741 and to change the form of output conversion. See section BF.11.01.

The conversion is implemented by the routines tty_read and tty_write, which are briefly described in BF.11.01.

Output

Output code conversion is very simple. Each ASCII character is treated in one of the following ways:

1. it is ignored.

2. it is translated into a single device character, possibly preceded by a case-shift.

3. it is translated into a "prefix code" appropriate to the device followed by another character (like red and black shift).

4. it is translated into an octal escape.

5. it is translated into a software escape sequence.

No other mapping is done: in particular, there is no replacement of tabs for spaces or vice versa. However, blanks are deleted if they follow a newline and are not followed by anything.

If more than a certain number of print-positions is generated, an overflow has occurred and a newline is inserted to prevent a pile-up at the right margin. The certain number is 74 for M37's, 124 for 1050's and 2741's. This number is changable via an I/O system order call. (BF.11.01) For purposes of calculating the line-length, as well as for character timing, tty_write assumes the standard tab settings of columns 11, 21, 31, ... There is no way to change tty_write's idea of these settings.

The treatment of characters depends on whether edited mode is in effect. Table I gives the character sequences generated for all the ASCII control characters and non-obvious graphics in normal mode. Table II does the same for edited mode. Non-ASCII characters (larger than octal 177) are escaped octally in normal mode, ignored in edited mode.

Edited mode conversion is not supported for 938 golf balls.

Input

The input code conversion module produces a partially canonical stream. The only important way in which canonicalization is imcomplete is that among the carriage-motion characters, only newline, space, backspace, and carriage return are handled properly; the others are treated like ordinary graphics. This means that horizontal and vertical tabs, forward and reverse half line feed, and reverse half line feed, and also red and black ribbon shift, come through just as they were typed. The canonicalization that is done means that

1. Characters appearing in the same print position are sorted in increasing ASCII collating sequence. This applies even if one of the characters was generated by an escape sequence.

2. The space character cannot appear in the same print position as another character.

3. The same character cannot appear twice in the same print position.

The 1050 and 2741 cannot print all the ASCII graphics. Certain characters on these devices are taken as stylized forms of ASCII characters. These are:

> ¢ for \
> ¬ for ~
> ! for '

With a 938 ball, the following characters are used:

> ¢ for \
> ¬ for ~
> ⌐ for |
> ! for '

Other unavailable characters must be escaped in.

The only escape character is the ASCII reverse slash (\), which is represented on the 1050 and 2741 by the cent sign (¢). Table III gives the defined escape sequences. The following rules also apply.

1. "\ddd", where each d̲ is an octal digit, generates the (possibly non-ASCII) character whose octal code is ddd. "\d" and "\dd" not followed by an octal digit generate 00d and 0dd respectively.

2. "\x" where x is not an octal digit or part of a defined escape sequence generates x itself. However, if x is a upper case letter it is mapped into the corresponding lower case letter. Since escape processing is done before erase and kill, this means that "\#" and "\@" generate "#" and "@" respectively. Also, "\\" generates "\", if the sequence is not immediately followed by "(" or ")".

3. To overstrike an escaped character, put the extra characters on top of the escape character. Thus, to get "[" on a 1050, type "¢<". The only way to overstrike two escape sequences is to escape in a backspace: to get "⌈" on a 1050, type "¢<¢010¢t".

The standard and only erase and kill characters are "#" and "@" respectively. (The inability to change these characters is mitigated by the ease of escaping them in.) The following rules relate to erase and kill processing.

1. "#" erases everything in its own print position and everything in the previous print position. Moreover, it erases all white space (blanks and tabs) immediately preceding it.

2. "@" erases everything in its own print position and everything in all preceding print positions. It does not necessarily erase everything typed before it if backspace or carriage return has been used.

3. Within a print position, erase processing is done before kill. This means that one can erase an "@" by typing "#" on

top of it. (Note that the previous print position is erased too.)

On an M37 Teletype, the prefix key initiates a sort of hardware escape sequence. Table IV gives the effect produced by following a prefix with each of several characters. If prefix is followed by any other character, both the prefix and the character are ignored.

The 1050 has a roughly analogous key, marked "altn coding". If this key and "8" are depressed simultaneously, the resulting prefix character and the next character are ignored. If this key and any other key are depressed simultaneously, the character may or may not be ignored.

A <u>Final</u> <u>Note</u>

With respect to the fine points of input conversion, a healthy spirit of experimentation is suggested.

ASCII Name	ASCII Code	Effect on 1050/2741	Effect on M37
NUL	000	¢000	\000
SOH	001	¢001	\001
STX	002	¢002	enter extra char. mode
ETX	003	¢003	leave extra char. mode
EOT	004	¢004	\004
ENQ	005	¢005	\005
ACK	006	turn on printer	turn on printer
BEL	007	¢007	sound bell
BS	010	backspace	backspace
HT	011	horizontal tab	horizontal tab
NL	012	newline	newline
VT	013	¢013	vertical tab
NP	014	¢014	new page
CR	015	¢015	carriage return
SO	016	red shift	red shift
SI	017	black shift	black shift
DLE	020	¢020	\020
DC1	021	¢021	\021
DC2	022	¢022	half line forward
DC3	023	¢023	full line reverse
DC4	024	¢024	half line reverse
NAK	025	printer off	printer off
SYN	026	¢026	horiz. tab set
ETB	027	¢027	horiz. tab clear
CAN	030	¢b	\b
EM	031	¢c	\c
SUB	032	¢032	\032
ESC	033	¢033	\033
FS	034	¢f	\f
GS	035	¢035	vertical tab set
RS	036	¢r	\r
US	037	¢037	vertical tab clear
'	047	'	'
<	074	< or ¢l	<
>	076	> or ¢g	>
[	133	¢< or ¢¢(	[
\	134	¢134	\134
]	135	¢> or ¢¢)	]
	136	or	`
`	140	¢'	
{	173	¢(	{
¦	174	¦ or	¦
}	175	¢)	}
~	176	¢t	~
DEL	177	ignored	ignored

Table I. Normal Output
The second alternative for
1050's and 2741's is for
a 938 ball.

ASCII Name	ASCII Code	Effect on 1050/2741	Effect on M37
NUL	000	ignored	ignored
SOH	001	ignored	ignored
STX	002	ignored	enter extra char. mode
ETX	003	ignored	leave extra char. mode
EOT	004	ignored	ignored
ENQ	005	ignored	ignored
ACK	006	turn on printer	turn on printer
BEL	007	ignored	sound bell
BS	010	backspace	backspace
HT	011	horizontal tab	horizontal tab
NL	012	newline	newline
VT	013	ignored	vertical tab
NP	014	ignored	new page
CR	015	ignored	carriage return
SO	016	red shift	red shift
SI	017	black shift	black shift
DLE	020	ignored	ignored
DC1	021	ignored	ignored
DC2	022	ignored	half line forward
DC3	023	ignored	full line reverse
DC4	024	ignored	half line reverse
NAK	025	printer off	printer off
SYN	026	ignored	horiz. tab set
ETB	027	ignored	horiz. tab clear
CAN	030	ignored	ignored
EM	031	ignored	ignored
SUB	032	ignored	ignored
ESC	033	ignored	ignored
FS	034	ignored	ignored
GS	035	ignored	vertical tab set
RS	036	ignored	ignored
US	037	ignored	vertical tab clear
'	047	'	'
<	074	<	<
>	076	>	>
[	133	←	[
\	134	¢	\
]	135	→	]
	136	¬	
`	140	blank	`
{	173	←	{
|	174	|	|
}	175	→	}
~	176	blank	~
DEL	177	ignored	ignored

Table II. Edited Output

Escape Sequence	Ascii character
\b, \B	CAN (030)
\c, \C	EM (031)
\f, \F	FS (035)
\g, \G	> (076)
\l, \L	< (074)
\r, \R	RS (035)
\t, \T	~ (176)
\<	[(133)
\>	] (135)
\'	` (140)
\(	{ (173)
\)	} (175)
\\(	[(133)
\\)	] (135)

Table III.   Escape sequences.

Typed	Generated
0	0 (060)
1	HT set (SYN, 026)
2	HT clear (ETB, 027)
3	RRS (SO, 016)
4	BRS (SI, 017)
5	VT set (GS, 035)
6	VT clear (US, 037)
7	FLR (DC4, 024)
8	HLR (DC3, 023)
9	HLF (DC2, 022)
:	P off (NAK, 025)
;	P on (ACK, 006)

Table IV.   Prefixed characters for M37