

Character Sets Used in Romanization

The transliterations produced by applying *ALA-LC Romanization Tables* are encoded in machine-readable form into USMARC records. Encoding of the basic Latin alphabet, special characters, and character modifiers listed in this publication is done in USMARC records following two American National Standards; the *Code for Information Interchange (ASCII)* (ANSI X3.4), and the *Extended Latin Alphabet Coded Character Set for Bibliographic Use (ANSEL)* (ANSI Z39.47). Each character is assigned a unique hexadecimal (base-16) code which identifies it unambiguously for computer processing.

The graphics for certain characters used in romanization are so similar that they are often confused. The apostrophe (’), soft sign (prime) (‘), acute accent (´), alif (ʿ), and ayn (ʿ) are typical of characters that are difficult to differentiate. A listing has been provided at the end of each romanization table that shows special characters and character modifiers that are used with it. The following list shows the entire basic Latin alphabet, and all special characters and character modifiers that are used in these romanization tables. Each unique character is presented with its graphic, hexadecimal code, and name. More information on the character sets used in USMARC records can be found in *USMARC Specifications for Record Structure, Character Sets, and Exchange Media* (Washington: Library of Congress, CDS, 1994).

Graphic	Hex Code	Name	Graphic	Hex Code	Name
	20	SPACE (BLANK)	?	3F	QUESTION MARK
!	21	EXCLAMATION MARK	@	40	COMMERCIAL AT
"	22	QUOTATION MARK	A	41	A
#	23	NUMBER SIGN	B	42	B
\$	24	DOLLAR SIGN	C	43	C
%	25	PERCENT SIGN	D	44	D
&	26	AMPERSAND	E	45	E
'	27	APOSTROPHE	F	46	F
(28	OPENING PARENTHESIS	G	47	G
)	29	CLOSING PARENTHESIS	H	48	H
*	2A	ASTERISK	I	49	I
+	2B	PLUS SIGN	J	4A	J
,	2C	COMMA	K	4B	K
-	2D	HYPHEN-MINUS	L	4C	L
.	2E	PERIOD (DECIMAL POINT)	M	4D	M
/	2F	SLASH	N	4E	N
0	30	0	O	4F	O
1	31	1	P	50	P
2	32	2	Q	51	Q
3	33	3	R	52	R
4	34	4	S	53	S
5	35	5	T	54	T
6	36	6	U	55	U
7	37	7	V	56	V
8	38	8	W	57	W
9	39	9	X	58	X
:	3A	COLON	Y	59	Y
;	3B	SEMICOLON	Z	5A	Z
<	3C	LESS THAN SIGN	[5B	OPENING SQUARE BRACKET
=	3D	EQUAL SIGN	\	5C	REVERSE SLASH
>	3E	GREATER THAN SIGN]	5D	CLOSING SQUARE BRACKET

Graphic	Hex Code	Name	Graphic	Hex Code	Name
\	5C	REVERSE SLASH	‘	B0	AYN
[5D	OPENING SQUARE BRACKET	Ÿ	B1	LOWERCASE POLISH L
^	5E	SPACING CIRCUMFLEX	ø	B2	LOWERCASE SCANDINAVIAN O
˘	5F	SPACING UNDERSCORE	đ	B3	LOWERCASE D WITH CROSSBAR
˘	60	SPACING GRAVE	þ	B4	LOWERCASE ICELANDIC THORN
a	61	a	æ	B5	LOWERCASE DIGRAPH AE
b	62	b	œ	B6	LOWERCASE DIGRAPH OE
c	63	c	“	B7	HARD SIGN (DOUBLE PRIME)
d	64	d	ı	B8	LOWERCASE TURKISH I
e	65	e	£	B9	BRITISH POUND SIGN
f	66	f	ð	BA	ETH
g	67	g	σ	BC	LOWERCASE O-HOOK
h	68	h	υ	BD	LOWERCASE U-HOOK
i	69	i	◊	CO	DEGREE SIGN
j	6A	j	ℓ	C1	LOWERCASE SCRIPT L
k	6B	k	Ⓟ	C2	PHONO COPYRIGHT MARK
l	6C	l	©	C3	COPYRIGHT MARK
m	6D	m	#	C4	SHARP
n	6E	n	¿	C5	INVERTED QUESTION MARK
o	6F	o	¡	C6	INVERTED EXCLAMATION MARK
p	70	p	¿	E0	PSEUDO QUESTION MARC
q	71	q	̀	E1	GRAVE
r	72	r	ˆ	E2	ACUTE
s	73	s	ˆ	E3	CIRCUMFLEX
t	74	t	ˆ	E4	TILDE
u	75	u	ˆ	E5	MACRON
v	76	v	ˆ	E6	BREVE
w	77	w	ˆ	E7	SUPERIOD DOT
x	78	x	ˆ	E8	UMLAUT (DIERESIS)
y	79	y	ˆ	E9	HACEK (CARON)
z	7A	z	ˆ	EA	CIRCLE ABOVE (ANGSTROM)
{	7B	OPENING CURLY BRACE	ˆ	EB	LIGATURE, FIRST HALF
	7C	VERTICAL BAR (FILL)	ˆ	EC	LIGATURE, SECOND HALF
}	7D	CLOSING CURLY BRACE	ˆ	ED	HIGH COMMA, OFF CENTER
~	7E	SPACING TILDE	ˆ	EE	DOUBLE ACUTE
Ł	A1	UPPERCASE POLISH L	ˆ	EF	CANDRABINDU
Ø	A2	UPPERCASE SCANDINAVIAN O	ˆ	F0	CEDILLA
Ð	A3	UPPERCASE D WITH CROSSBAR	ˆ	F1	RIGHT HOOK (OGONEK)
Þ	A4	UPPERCASE ICELANDIC THORN	ˆ	F2	DOT BELOW
Æ	A5	UPPERCASE DIGRAPH AE	ˆ	F3	DOUBLE DOT BELOW
Œ	A6	UPPERCASE DIGRAPH OE	ˆ	F4	CIRCLE BELOW
/	A7	SOFT SIGN (PRIME)	ˆ	F5	DOUBLE UNDERSCORE
·	A8	DOT IN MIDDLE OF LINE	ˆ	F6	UNDERSCORE
ˆ	A9	MUSICAL FLAT	ˆ	F7	LEFT HOOK (COMMA BELOW)
®	AA	SUBSCRIPT PATENT MARK	ˆ	F8	RIGHT CEDILLA
±	AB	PLUS OR MINUS	ˆ	F9	UPADHMANTIYA
Œ	AC	UPPERCASE O-HOOK	ˆ	FA	DOUBLE TILDE, FIRST HALF
Ū	AD	UPPERCASE U-HOOK	ˆ	FB	DOUBLE TILDE, SECOND HALF
’	AE	ALIF	ˆ	FE	HIGH COMMA, CENTERED