

GEDCOM ANSEL

The GEDCOM specification describes a character encoding scheme referred to as ANSEL which is an extended version of ANSEL ANSI Z39.47-1985. This document describes code point to character mapping for a file whose bytes represent characters encoded using this customised character encoding scheme.

UNICODE code points are provided to facilitate the conversion of files to UNICODE.

ANSEL ANSI Z39.47-1985

Code points below 80 should be mapped to the corresponding character defined in the US ASCII character encoding scheme. The following tables represent code point to character mapping for all other code points with defined characters.

ANSEL Nonspacing Graphic Characters

ANSEL code point (hex)	UNICODE code point (hex)	Description
E0	0309	low rising tone mark (◌̈́)
E1	0300	grave accent (◌̀)
E2	0301	acute accent (◌́)
E3	0302	circumflex accent (◌̂)
E4	0303	tilde (◌̃)
E5	0304	macron (◌̄)
E6	0306	breve (◌̆)
E7	0307	dot above (◌̈)
E8	0308	umlaut (dieresis) (◌̈)
E9	030C	hacek (◌̣)
EA	030A	circle above (angstrom) (◌̊)
EB	FE20	ligature, left half (◌̸)
EC	FE21	ligature, right half (◌̶)
ED	0315	high comma, off center (◌̑)
EE	030B	double acute accent (◌̂́)
EF	0310	candrabindu (◌̃̄)
F0	0327	cedilla (◌̣)
F1	0328	right hook (◌̑)
F2	0323	dot below (◌̣)
F3	0324	double dot below (◌̣̣)
F4	0325	circle below (◌̈́)
F5	0333	double underscore (◌̣̣)
F6	0332	underscore (◌̣)
F7	0326	left hook (◌̑)
F8	031C	right cedilla (◌̣)
F9	032E	half circle below (upadhmaniya) (◌̈́)
FA	FE22	double tilde, left half (◌̸)
FB	FE23	double tilde, right half (◌̶)
FE	0313	high comma, centered (◌̑)

ANSEL Spacing Graphic Characters

ANSEL code point (hex)	UNICODE code point (hex)	Description
A1	0141	slash L - uppercase (Ł)
A2	00D8	slash O - uppercase (Ø)
A3	0110	slash D - uppercase (Ð)
A4	00DE	thorn - uppercase (Þ)
A5	00C6	ligature AE - uppercase (Æ)
A6	0152	ligature OE - uppercase (Œ)
A7	02B9	miagkii znak ('))
A8	00B7	middle dot (·)
A9	266D	musical flat (b)
AA	00AE	patent mark (®)
AB	00B1	plus or minus (±)
AC	01A0	hook O - uppercase (Ŏ)
AD	01AF	hook U - uppercase (Ū)
AE	02BC	alif (ʿ)
B0	02BB	ayn (ع)
B1	0142	slash l - lowercase (ł)
B2	00F8	slash o - lowercase (ø)
B3	0111	slash d - lowercase (ð)
B4	00FE	thorn - lowercase (þ)
B5	00E6	ligature ae - lowercase (æ)
B6	0153	ligature oe - lowercase (œ)
B7	02BA	tverdyj znak (")
B8	0131	dotless i - lowercase (ı)
B9	00A3	British pound (£)
BA	00F0	eth (ð)
BC	01A1	hook o - lowercase (ơ)
BD	01B0	hook u - lowercase (ư)
C0	00B0	degree sign (°)
C1	2113	script l (ℓ)
C2	2117	phonograph record copyright mark (℗)
C3	00A9	copyright symbol (©)
C4	266F	musical sharp (#)
C5	00BF	inverted question mark (¿)
C6	00A1	inverted exclamation mark (¡)

GEDCOM extensions of ANSEL ANSI Z39.47-1985

Different GEDCOM specifications define up to 6 extension characters which are not part of ANSEL ANSI Z39.47-1985.

ANSEL code point (hex)	UNICODE code point (hex)	Description	GEDCOM versions
BE	25A1	empty box (□)	5.3
BF	25A0	black box (■)	5.3
CD	0065	e in middle of line (e)	5.3
CE	006F	o in middle of line (o)	5.3
CF	00DF	es zet (ß)	5.3, 5.4, 5.5, 5.5.1, 5.6
FC	0338	diacritic slash through char (⸗)	5.3

Copyright

This document is made available under the following license:

Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

