

## 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 0x80 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
0xE0	0x0309	low rising tone mark (◌̈́)
0xE1	0x0300	grave accent (◌̀)
0xE2	0x0301	acute accent (◌́)
0xE3	0x0302	circumflex accent (◌̂)
0xE4	0x0303	tilde (◌̃)
0xE5	0x0304	macron (◌̄)
0xE6	0x0306	breve (◌̆)
0xE7	0x0307	dot above (◌̈)
0xE8	0x0308	umlaut (dieresis) (◌̈)
0xE9	0x030C	hacek (◌̣)
0xEA	0x030A	circle above (angstrom) (◌̊)
0xEB	0xFE20	ligature, left half (◌̯)
0xEC	0xFE21	ligature, right half (◌̰)
0xED	0x0315	high comma, off center (◌̑)
0xEE	0x030B	double acute accent (◌̇)
0xEF	0x0310	candrabindu (◌̐)
0xF0	0x0327	cedilla (◌̸)
0xF1	0x0328	right hook (◌̡)
0xF2	0x0323	dot below (◌̣)
0xF3	0x0324	double dot below (◌̣̣)
0xF4	0x0325	circle below (◌̤)
0xF5	0x0333	double underscore (◌̨)
0xF6	0x0332	underscore (◌̨)
0xF7	0x0326	left hook (◌̢)
0xF8	0x031C	right cedilla (◌̥)
0xF9	0x032E	half circle below (upadhmaniya) (◌̦)
0xFA	0xFE22	double tilde, left half (◌̰)
0xFB	0xFE23	double tilde, right half (◌̱)
0xFE	0x0313	high comma, centered (◌̑)

## ANSEL Spacing Graphic Characters

ANSEL code point (hex)	UNICODE code point (hex)	Description
0xA1	0x0141	slash L - uppercase (Ł)
0xA2	0x00D8	slash O - uppercase (Ø)
0xA3	0x0110	slash D - uppercase (Ð)
0xA4	0x00DE	thorn - uppercase (Þ)
0xA5	0x00C6	ligature AE - uppercase (Æ)
0xA6	0x0152	ligature OE - uppercase (Œ)
0xA7	0x02B9	miagkii znak ('))
0xA8	0x00B7	middle dot (·)
0xA9	0x266D	musical flat (b)
0xAA	0x00AE	patent mark (®)
0xAB	0x00B1	plus or minus (±)
0xAC	0x01A0	hook O - uppercase (Ŏ)
0xAD	0x01AF	hook U - uppercase (Ū)
0xAE	0x02BC	alif (ʿ)
0xB0	0x02BB	ayn (ع)
0xB1	0x0142	slash l - lowercase (ł)
0xB2	0x00F8	slash o - lowercase (ø)
0xB3	0x0111	slash d - lowercase (ð)
0xB4	0x00FE	thorn - lowercase (þ)
0xB5	0x00E6	ligature ae - lowercase (æ)
0xB6	0x0153	ligature oe - lowercase (œ)
0xB7	0x02BA	tverdyj znak (")
0xB8	0x0131	dotless i - lowercase (ı)
0xB9	0x00A3	British pound (£)
0xBA	0x00F0	eth (ð)
0xBC	0x01A1	hook o - lowercase (ơ)
0xBD	0x01B0	hook u - lowercase (ư)
0xC0	0x00B0	degree sign (°)
0xC1	0x2113	script l (ℓ)
0xC2	0x2117	phonograph record copyright mark (℗)
0xC3	0x00A9	copyright symbol (©)
0xC4	0x266F	musical sharp (#)
0xC5	0x00BF	inverted question mark (¿)
0xC6	0x00A1	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
0xBE	0x25A1	empty box (□)	5.3
0xBF	0x25A0	black box (■)	5.3
0xCD	0x0065	e in middle of line (e)	5.3
0xCE	0x006F	o in middle of line (o)	5.3
0xCF	0x00DF	es zet (ß)	5.3, 5.4, 5.5, 5.5.1, 5.6
0xFC	0x0338	diacritic slash through char (ø)	5.3

### Copyright

This document is made available under a Creative Commons Attribution-ShareAlike 3.0 Unported License.