

# Unicode and UTF-8 Support

UltraEdit provides support for Unicode (16-Bit wide character) files, and for UTF-8 files and allows direct editing of Unicode files and UTF-8 files as well as conversion routines between ASCII/ANSI and Unicode or UTF-8. UltraEdit attempts to detect the file type when the file is loaded. It will look for the FF FE marker for Unicode files. For UTF-8 it will look for one of the following three occurrences:

- 1) File marker ( BOM) EF, BB, BF.
- 2) String occurrences "charset=utf-8" or "encoding=utf-8"
- 3) It will look for the occurrence of valid UTF-8 multi-byte characters in the first 64KB of a file.

If the file is found to be Unicode, it will be treated as such and the status bar will indicate this with "U-" being displayed in front of the file type in the status bar.

If the file is found to be UTF-8, it will be treated as such and converted internally to Unicode (16-Bit) for editing. The status bar will indicate this with "U8-" being displayed in front of the file type in the status bar. When the file is saved, it will be converted back from Unicode to UTF-8 and saved in this format.

Handling internally as Unicode allows the Unicode character messages from the OS to be used for character input. Additionally, the WM\_UNICHAR message used in some versions of Windows and with some applications is supported by UltraEdit allowing Unicode characters to be entered on versions of the OS that do not support Unicode natively (Windows 95/98/Me).

The following conversions to and from Unicode/UTF-8 are available:

ASCII to Unicode	converts from ASCII to Unicode
UTF-8 to Unicode	converts from UTF-8 to Unicode (16-Bit)
Unicode to ASCII	converts from Unicode to ASCII
UTF-8 to ASCII	converts from UTF-8 to ASCII
ASCII to UTF-8 (Unicode Editing)	converts from ASCII to UTF-8 with the file internally in UNICODE format for editing

Unicode/UTF-8 to UTF-8 (Unicode)	converts the file from either Unicode or UTF-8 (non Unicode internally) to UTF-8 with the file internally in UNICODE format for editing
Unicode/ASCII/UTF-8 to UTF-8 (ASCII)	converts from Unicode, ASCII or UTF-8 (Unicode format internally) to UTF-8 but leaves the file in non-unicode (ASCII display) format for editing.

Unicode file support within UltraEdit is provided on Windows 95, 98, Me, NT and 2000.

As not all operating systems natively support Unicode certain functions convert from Multi-Byte to Unicode or from Unicode to Multi-Byte. If the active code page (see Windows documentation/help) does not support the characters then the conversion will not convert correctly.

The functions that convert between Multi-Byte and Unicode include Find/Replace and Column Insert.

Posted - Thu, Oct 13, 2011 6:11 PM. This article has been viewed 2614 times.

Online URL: <http://www.ultraedit.com/help/article/unicode-and-utf-8-support-1248.html>

---