#¹$²K³UConvert -- Convert to/from unicode text.

The Unicode Converter can be used to convert a file to Unicode from an ANSI, OEM, or other type of code page. You can also use the Conversion menu to bring up dialogs which allow you to specify conversion options.

To use the Unicode Converter:

1   From the File menu, choose Open Source File, and then specify the file you want to convert.

2   From the Conversion menu, choose Convert now.

3   In the Save As dialog box, specify a filename for the converted file.

More information is available on:

Clipboard support.

Installing conversion tables.

Conversion options.

Byte order.

Viewing text data.

#[4]$[5]K[6]Clipboard support

UConvert will accept text from the clipboard as source data.   It will also copy destination data back to the clipboard.   Use the Edit menu, or traditional accelerator keys, to access these features.

Clipboard text data is automatically tagged by the Windows system with respect to the character encoding that it uses.

#[7]$[8]K[9]Installing conversion tables

UConvert also allows you to install additional conversion tables without using REGEDT32.EXE.   To do this, choose the Install New Conversion Tables command from the Conversion menu, and choose the Add button in the Conversion Tables dialog box so that you can select additional .NLS files to add to the conversion list. The .NLS files provided with Windows NT are stored in the *SystemRoot*\SYSTEM32 directory.

#[10]$[11]K[12]Conversion options

When converting **MultiByteToWideChar**:

When MB_PRECOMPOSED is specified, the program always uses precomposed characters--that is, characters in which a base character and a nonspacing character have a single character value. This is the default translation option.

When MB_COMPOSITE is specified, the program uses composite characters--that is, characters in which a base character and a nonspacing character have different values.

When MB_USEGLYPHCHARS is specified, the program uses glyph characters instead of control characters.

(A base character is a character followed by a nonspacing character. For example, in an accented e, the e is the base character and the accent mark is the nonspacing character.)

When converting **WideCharToMultiByte**:

When WC_COMPOSITECHECK is specified, the program converts composite characters to precomposed characters. In composite characters, the base character and the corresponding nonspacing character have different character values. In precomposed characters, the base character and the corresponding nonspacing character have a single character value.

When an application specifies WC_COMPOSITECHECK, it can use the following flags to customize the conversion to precomposed characters. These flags determine the function's behavior when there is no precomposed mapping for a base-nonspace character combination in the wide-character string.

WC_DISCARDNS	Discard nonspacing characters.
WC_SEPCHARS	Generate separate characters. This is the default behavior.
WC_DEFAULTCHAR	Replace exceptions with the default character.

#¹³$¹⁴K¹⁵Byte order

When reading a source file, UConvert will look for the Unicode *byte order mark*.   If it locates this particular two byte sequence at the beginning of a file, it will assume that the remainder of the text is a proper sequence of wide characters.   UConvert will also look for the Unicode *reverse byte order mark*.   In this case, it will again assume Unicode text, however, it will automatically swap the low and high bytes of every wide character.

At any time, the user may swap the byte order of the source or destination by using Swap byte order menu items under the Conversion menu.   If this action is taken, the main window will display **Byte Order Reversed.**

Byte order is important only when moving between systems which run in different *endian* modes.   For example, Intel processors all run *little endian*, while Motorola processors run *big endian*.   As of the version 3.1 release of Windows NT, it is running only on little endian processors.   However, Unicode is supported on a variety of different systems.

---

¹³# Order

---

¹⁴$ Byte order

---

¹⁵K Byte order

#¹⁶$¹⁷K¹⁸Viewing text data

There are two buttons on the main UConvert window which allow the user to view the source or destination text data.   When viewing textual data, it is necessary to have a font which provides *glyphs* to display the characters in the text.   The UConvert program will attempt to select an appropriate font from the ones currently installed on the system.   However, UConvert can handle a larger set of code pages than there are typically fonts installed for.   In this case, the font system will offer the best match possible, and characters not supported by the font will be displayed using a special default character.   For many fonts, this default character is a box.

Windows NT version 3.1 was released with a special font named **Lucida Sans Unicode** which is contained in the file **L_10646.ttf**.   By installing this font through the control panel, the user will be able to see a wider range of characters when attempting to view Unicode data.

₁₆**# Viewing**

₁₇**$ Viewing text data**

₁₈**K Viewing text data**