

Chapter III-16 — Text Encodings

When a given text file is successfully converted, its entry in the Text Encoding column changes to "UTF-8".

After doing a conversion, the Convert Text Files button changes to Refresh. Clicking Refresh refreshes the list, showing text files remaining to be converted, if any.

Text Encoding Names and Codes

Igor operations and functions such as `ConvertTextEncoding`, `SaveNotebook`, `WaveTextEncoding` and `SetWaveTextEncoding` accept text encoding codes as parameters or return them as results.

The functions `TextEncodingName` and `TextEncodingCode` provide conversion between names and codes. For most Igor programming purpose, you need the code, not the name.

For each text encoding supported by Igor there is one text encoding code. This code may correspond to one or more text encoding names. For example, code 2 is the Macintosh Roman text encoding and corresponds to the text encoding names "macintosh", "MacRoman" and "x-macroman".

Because spelling of text encoding names is inconsistent in practice, Igor recognizes variant spellings such as "Shift JIS", "ShiftJIS", "Shift_JIS" and "Shift-JIS". When comparing text encoding names, Igor ignores all non-alphanumeric characters. It also ignores leading zeros in numbers embedded in text encoding names so that "ISO-8859-1" and "ISO-8859-01" refer to the same text encoding. It uses a case-insensitive comparison.

Some of the entries in the table below are marked NOT SUPPORTED. These are not supported because the ICU (International Components for Unicode) library, which Igor uses for text encoding conversions, does not support it. "Not supported" means that the `TextEncodingName` and `TextEncodingCode` functions do not recognize these text encodings and Igor can not convert to them or from them.

Text Encoding Name	Code	Notes
None	0	Means the text encoding is unknown
UTF-8	1	Unicode UTF-8
macintosh, MacRoman, x-macroman	2	Macintosh Western European
Windows-1252	3	Windows Western European
Shift_JIS	4	Predominant Japanese text encoding
MacJapanese, x-mac-japanese	4	Virtually the same as Shift_JIS
Windows-932	4	Virtually the same as Shift_JIS
EUC-JP	5	Japanese, typically used on Unix
Big5	20	Traditional Chinese
x-mac-chinesetrad	20	Virtually the same as Big5
Windows-950	20	Virtually the same as Big5
EUC-CN	21	Simplified Chinese
x-mac-chinesesimp	21	Simplified Chinese
Windows-936	21	Simplified Chinese
ISO-2022-CN	22	Simplified Chinese
GB18030	23	Official text encoding of the PRC. Encompasses Traditional Chinese and Simplified Chinese. Compatible with Windows-936.
EUC-KR, x-mac-korean	40	Macintosh Korean