

Chapter III-16 — Text Encodings

the BOM if you later save the file. This process of reloading an externally-modified plain text file overwrites the writeBOM flag that you may have set using the Text Encoding dialog or Notebook operation.

BOMs are irrelevant for formatted text notebooks.

The **FReadLine** operation looks for and skips UTF-8 byte order marks and UTF-16 byte order marks if you specify UTF-16 using the /ENCG flag.

Formatted Text Notebook File Text Encodings

Formatted notebook files use an Igor-specific file format. You can open such a file as a formatted notebook or as an Igor help file.

Unlike plain text files, in which all text must be encoded using a single text encoding, Igor6 formatted notebook files can use multiple text encodings in a single file. In Igor6, the text encoding for a given format run is determined by the font controlling that run. For example, a single Igor6 formatted text file can contain western text and Shift JIS (Japanese) text encodings. In fact a single paragraph can contain multiple text encodings.

We use the term MBCS (multi-byte character system) to distinguish these Igor6 formatted text files from formatted text files saved using UTF-8 by Igor7 or later.

Igor now stores text internally (in memory) as UTF-8. Consequently it must convert MBCS text into UTF-8 when opening Igor6 formatted files. When doing this conversion, Igor uses text encoding information written to the file by IP6.3x and later. Pre-IP6.3x formatted text notebook files lack this text encoding information and so may be misinterpreted. If you open a formatted text notebook and you get incorrect characters or text encoding conversion errors, try opening and resaving the file in Igor Pro 6.3x, and then reopen it.

In Igor Pro 8.00 or later, Igor writes new formatted notebooks and formatted notebooks previously written using UTF-8 as UTF-8. It writes in Igor6-compatible MBCS format only when saving a file that was previously written as MBCS and only if all of the file's characters can be represented in MBCS. If you open a UTF-8 formatted notebook file in Igor6, any non-ASCII characters will appear garbled.

To determine how a particular file is saved, click the page icon in the lower-lefthand corner of a formatted notebook or help window. The resulting File Information dialog will say either "MBCS" or "UTF-8" in the Text Encoding area.

If you open a UTF-8-formatted text notebook in Igor Pro 6.36 or before and then save the notebook, it is marked as an Igor6 notebook using MBCS even though it really contains UTF-8 text. If you subsequently open the notebook in Igor7 or later, all non-ASCII characters will be incorrectly interpreted as MBCS and will be wrong.

If you open a UTF-8-formatted text notebook in Igor Pro 6.37 or a later 6.x version and then save the notebook, it is marked as using UTF-8 text encoding. Igor6 will still display the wrong characters for non-ASCII text, but if you subsequently open the notebook in Igor7 or later, the non-ASCII characters will be correctly interpreted as UTF-8.

Wave Text Encodings

Each wave internally stores several plain text elements. They are:

- wave name
- wave units
- wave note
- wave dimension labels
- text wave content (applies to text waves only)