

# How can I get `WideCharToMultiByte` to convert strings encoded in UTF-16BE?

 [devblogs.microsoft.com/oldnewthing/20231005-00](https://devblogs.microsoft.com/oldnewthing/20231005-00)

October 5, 2023



Raymond Chen

A customer had a Windows program that receives data in UTF-16BE format, and they want to convert it to Shift JIS format. According to the customer liaison:

They convert the characters from UTF-16LE to Shift JIS by calling `WideCharToMultiByte`, and it works fine. However, trying to convert the characters from UTF-16BE to Shift JIS via `WideCharToMultiByte` produces garbage. How can we tell `WideCharToMultiByte` that the string is UTF-16BE? Is there any documentation that explains this?

In Windows, if a string is described as being in Unicode or UTF-16 format, the documentation means UTF-16LE format by default. Similarly, if a sequence of bytes is described as encoding a multi-byte integer, the documentation means little-endian twos-complement format by default.<sup>1</sup>

The bias toward little-endian format in Windows is so strong that big-endian format is sometimes called “reverse byte order”, such as in the values returned by the `IsTextUnicode` format.

In this case, it’s not clear how the customer is using the `WideCharToMultiByte` function to convert UTF-16BE to Shift JIS. The `WideCharToMultiByte` function does not have any flag to specify the source encoding, so the system assumes the default, which is UTF-16LE. I’m guessing that they are just passing UTF-16BE data directly to the `WideCharToMultiByte` function and hoping that the function somehow employs psychic powers to realize “Oh, this time, the data should be treated as UTF-16BE.”

The `WideCharToMultiByte` function does not have psychic powers. It converts from UTF-16LE.

The customer must convert their source data from UTF-16BE to UTF-16LE, and then pass the UTF-16LE data to `WideCharToMultiByte` function. Fortunately, converting UTF-16BE to UTF-16LE is extremely straightforward.

<sup>1</sup> One example of how the default might not apply is when talking about data encoded in “network byte order”.