Subject: Re: 16 bits wchar
Posted by cbpporter on Wed, 03 Oct 2007 12:43:21 GMT
View Forum Message <> Reply to Message

Quote:
However, please check the fixed version Stream::GetUtf8():

Thank you! You should have said that you would fix it that quickly and I wouldn't have tried it myself . Shouldn't second if clause be < C2?

Quote:
What is the point of spreading encoding related stuff all over the application? Stream works with bytes, end of story. I do not want to end with multiple methods for everything that can handle text.

Yes, I agree, Stream should work with bytes. But text processing should never work with bytes, unless in legacy mode.

And considering the problem regarding escaping, AFAIK, if the sixth byte is invalid, you need to signal an error for the first byte and continue to decode the second character as a new code point.

And also six byte Utf-8 is no longer considered correct, and should only be used when legacy data needs to be processed. But since 4 bytes allow well over 1 million code-units, I doubt there is any data stored in six bytes format. CESU8 is another thing though, but that is not supported, so it's not a problem.