
Subject: Choosing the best way to go full UNICODE
Posted by [mirek](#) on Sat, 27 May 2017 14:58:08 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hi,

last week, for the first time, I have encountered a problem caused by presence of characters outside of UCS2 range (did you know that emoji characters have assigned codepoints? I did not either. Yet our www backend was failing because somebody send "thumbs up" in the message...

So perhaps it is time for full support. The mission target is that these characters are correctly loaded/stored to/from LineEdit, DocEdit, EditString, RichText.

Now I can see 3 paths to achieve this:

(1) Extend WString to 32-bits (so that sizeof(wchar) is 4 bytes). This is the most straightforward and right now my favorite. Downsides is that I am a little bit afraid about backward compatibility and that the performance will be worse than WString. Performance is less of an issue because I believe that in most cases, texts are stored as UTF-8 String anyway, but worth checking.

(2) Add 32-bit LString (or "UString"?) and perhaps deprecate WString. Fixes the problem with compatibility, but fattens the API.

(3) Keep WString and somehow fix things to work with surrogate pairs.

Any thoughts?

Mirek
