Subject: Re: Will UPP support full UNICODE (21bits long codepoint)?
Posted by mirek on Mon, 17 Aug 2020 09:01:35 GMT
View Forum Message <> Reply to Message

Oblivion wrote on Sat, 15 August 2020 13:15Quote: But I am not at the moment sure whether combining characters are the only source of multi-codepoint graphemes.

Yeah, there are at least surrogate pairs (Since U++ use 16-bit wchar), ligatures and IIRC some hangul graphemes. Anything else that I miss?

Surrogate pairs are rather well formed, but ligatures and multi-codepoint CJK/Devanagari stuff may pose problems...

Edit: Ah yes, what I miss is explained under the grapheme clusters section: http://www.unicode.org/reports/tr29/#Grapheme_Cluster_Bounda ries

I guess this might be the path forward:

https://en.wikipedia.org/wiki/HarfBuzz

It looks like most toolkits simply use HarfBuzz anyway... :)

Mirek