Subject: Re: Choosing the best way to go full UNICODE
Posted by cbpporter on Wed, 14 Jun 2017 10:42:01 GMT

View Forum Message <> Reply to Message

mirek wrote on Wed, 14 June 2017 13:30Good find. What do you suggest to do about that?

- I can leave current code as is and perhaps add "FullDecompose" variant.
- I can decompose it to 3 codepoints outright

I think I like the second option better...

Here is what I am implementing right now:
1. A Decompose method. You give it a code point and it gives you raw UnicodeData.txt data. That problematic character will still give you two results. This is already done.

But I practice I doubt it will ever be used, so much so that it barely qualifies as a public method. Instead, everybody will use...

2. A ToNFD() method. NFD is the canonical decomposition. The problematic character will result in 3 code points. This will be the main public method.

So my main method will be the method you prefer, giving 3 results. I just gave it the Unicode name.