

---

Subject: Re: String.GetCount with umlaut  
Posted by [mr\\_ped](#) on Wed, 07 Sep 2011 07:47:57 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

UTF8 is multibyte encoding with variable length for different characters.

IMHO the most correct solution for you is to convert from UTF8 to WString (UCS2), which should be well enough to cover \*all\* latin-like alphabets, Russian's azbuka as well and some more. (I'm afraid UCS2 does not cover all Chinese characters/etc.. no time to google ). Then work with WString to count characters, etc..

edit: Mirek beat me...

Quote:(Anyway, thinking about it, perhaps I shall add "GetCharCount()" method to String...).

Indeed for sure, actually UTF8 would love to have full set of functions like len/right/left/mid/etc.. whenever number of characters (position) is used as parameter. Plus special section in documentation to explain it.

---