Subject: Re: 16 bits wchar
Posted by mirek on Wed, 26 Sep 2007 20:40:29 GMT

cbpporter wrote on Wed, 26 September 2007 01:43sergei wrote on Wed, 26 September 2007 01:56
I didn't mention that I tested basic read/write performance. UTF handling would add overhead to 8 and 16 formats, but not to 32 format. I also remembered the UTF8-EE issue. UTF-32 could solve it easily. IIRC only 21 bits are needed for full unicode, so there's plenty of space to escape to (without overtaking private space).

The only problem with UTF-32 is the storage space. It is 2/4 times the size of UTF-8 and almost always double of UTF-16. And I don't think that UTF-8EE is such a big issue, you just have to make sure to use a more permissive validation scheme. And what is RTL anyway?

Not necessary. Current way of handling with this is just everything is mass stored as UTF-8 and only converted to UCS-2 for processing.

I guess this system should stand.

The only real trouble (and the main reason why sizeof(wchar) is 2) is Win32 compatibility. I do not feel well converting every text to UTF-16 for displaying on the screen... while, in reality, for 99% applications UCS-2 is enough...

Mirek