Subject: Re: 16 bits wchar
Posted by cbpporter on Wed, 06 Aug 2008 11:33:03 GMT
View Forum Message <> Reply to Message

Well, I was going to go life Today with my changes. I was issue free for a couple of days and I wanted to merge everything into Core on my local setup to see if there are any issues that I haven't discovered yet.

But I read something interesting. It seems that Unicode 5.1 gives quite specific information regarding what to do with ill formated texts, and most importantly, where the boundaries of such text are. In 5.0, there was a lot of place for interpretation, and pragmatically speaking is a good change.

To quote Unicode:
Quote:A process which interprets a Unicode string must not interpret any ill-formed code unit subsequences in the string as characters. (See conformance clause C10.) Furthermore, such a process must not treat any adjacent well-formed code unit sequences as being part of those ill-formed code unit sequences.
This does change slightly the result of error escaping in some cases. I don't know if it is important or not. I'll have to think about it.