
Subject: Re: JavaScriptCore

Posted by [cbpporter](#) on Thu, 12 Aug 2010 07:55:36 GMT

[View Forum Message](#) <> [Reply to Message](#)

I have investigated further the problem of ToUpper. After looking over the tables we have in U++, they have even better coverage than I thought. Those unit tests that failed must have been extremely thorough to have reported such a failure. So I figured that maybe not the limited coverage of the 2048 characters is the problem, rather maybe the table has some errors.

And indeed, there are some errors.

with stroke".

After some superficial testing, running the current ToUpper/ToLower on the whole 65536 range I have found 568/569 errors. This is really great news, seeing as the table only covers 2048 characters. This means that most of Unicode is case agnostic and we can get away with good support without using huge tables. Running it on only 2048 characters, I have found 50/43. I hope I used the correct testing method.

So if somebody can point me to the bit setup of uni__info, I could correct the values for the first 2048 characters. I can figure out most, but it would be better if I could get the real layout of that packed bitfield. If there are any free bits left, I have some data that I would like to store there, like if the character is punctuation, if it is Latin, etc.
