

---

Subject: Re: Choosing the best way to go full UNICODE

Posted by [cbpporter](#) on Tue, 06 Jun 2017 11:39:19 GMT

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

---

mirek wrote on Tue, 06 June 2017 14:21cbpporter wrote on Tue, 06 June 2017 11:18  
Characters are uniquely defined, so are the canonical composition and decomposition rules,  
together with compatibility substitutions.

I suppose so. Still learning.

Your lib is BSD? Available somewhere?

Oh yeah, the Unicode spec is huge. But it also recommends that you implement as much as you  
need for your needs, not the whole thing.

The lib is a bit more complicated. It is Apache Version 2.0, but not really released yet. More  
precisely, the more advanced Unicode parts only exist on my disks yet, they are not committed.  
But when they will be committed, it will be under Apache. How does future licensing work? :)

But I can share that, with the only caveat that the only things I'm really doing and are meant for  
the final lib is conversion and encoding of UnicodeData.txt. I'm not planning on composition  
handling for now since it is not about GUI or displaying text. My encoding scheme handles 4  
things: category, upper, lower and title case.

I would like still to add script to that data and probably you can't dodge forever adding canonical  
normalization support.

---