

---

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

Posted by [mirek](#) on Mon, 12 Jun 2017 08:13:44 GMT

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

---

cbpporter wrote on Mon, 12 June 2017 09:39What? How? Need to check it out.

I made a mistake in the estimate and actually how much data I need and the 9k got up to 34. So I started working on a 3 table solution :).

Really trivial. 4 Vectors of dwords (original code, 3 decomposed codes), then delta it (that will result in most values being the same), then ZCompress... (that will RLE and Huffman those same values).

That said, I am only doing decomposition of characters in UnicodeData.txt.

Also, I perhaps need to add this

[https://en.wikipedia.org/wiki/Korean\\_language\\_and\\_computers# Hangul\\_in\\_Unicode](https://en.wikipedia.org/wiki/Korean_language_and_computers#Hangul_in_Unicode)

too...

---