Appendix D on Variants

D.3 Base Character Evaluation

D.3.1 Caron vs Hook

Code Points Analyzed:

|  |  |  |
| --- | --- | --- |
| **Code Points** | **Glyph** | **Name** |
| 010F | ď | Letter D with Caron |
| 0257 | ɗ | Letter D with Hook |

Sequence D with Caron vs D with hook compared using Google Fonts in <https://wordmark.it/> :



Findings:

* Variant – indistinguishable, depending on font design.

D.3.2 K vs K with Hook

* Code Points Analyzed:

Sequence K (006B) and K with hook: ƙ (0199) compared using Google Fonts in <https://wordmark.it/> :



Findings:

* Variant – indistinguishable in some fonts

D.3.1 I vs Dotless I vs Iota

Code Points Analyzed:

|  |  |  |
| --- | --- | --- |
| 69 | i | Latin Small Letter I |
| 131 | ı | Latin Small Letter Dotless I |
| 269 | ɩ | Latin Small Letter Iota |

Sequence I vs Dotless I compared using Google Fonts in <https://wordmark.it/> :



Sequence Dotless I vs Iota compared using Google Fonts in <https://wordmark.it/> :



Findings for I vs Dotless I:

* The dot is readily perceptible . . . IF the user is sensitized by familiarity to the possibility of a dotless I (which only occurs in a couple of languages). Otherwise, the user sees what he expects to see. But based on the criteria for visual similarity, this would not be a variant.

Findings for Dotless I vs Iota:

* In the italic versions of any of the serif fonts (e.g. Times New Roman or Consolas) these are identical. For example:

*ı vs ɩ (Times New Roman) ı vs ɩ (Consolas)*

Therefore – Variant

Additional Findings:

* Mats made the case, based on case folding, for I and Dotless I being treated as variants. I concur, for this and other reasons. But this is a perception argument, not an argument based on basic letter shape one which is considered here.

D.4 Diacritics Stacking Evaluation

D.4.1 Stacking in Courier New (and perhaps other fonts)

We have seen that, with precomposed code points, there is no stacking problem. However, when we have not had a precomposed code points available, we have necessarily used combining diacritics. Then, the situation changes. In particular, when using the Courier New font (which is one of our three standard fonts for analysis), there is sometimes a problem. Sometimes, the combining mark simply gets its own space, with the following letter shifter right to make room – which is irritating, but not confusing. However in other cases the combining mark appears to be associated with the *following* letter.

Code Points Analyzed:

|  |  |  |
| --- | --- | --- |
| 1EB9 + 0301 | ẹ́ | Latin Small Letter E with Dot Below + Combining Acute Accent |
| 1EB9 + 0300 | ẹ̀ | Latin Small Letter E with Dot Below + Combining Grave Accent |
| 0067 + 0303 | g̃ | Latin Small Letter G + Combining Tilde |
| 0268 + 0303 | ɨ̃ | Latin Small Letter I with Stroke + Combining Tilde |
| 1ECD + 0300 | ọ̀ | Latin Small Letter O with Dot Below + Combining Grave Accent  |
| 1ECD + 0301 | ọ́ | Latin Small Letter O with Dot Below + Combining Acute Accent |
| 025B + 0331 + 0308 | ɛ̱̈ | Latin Small Letter Open E + Combining Diaresis + Combining Macron Below |
| 025B + 0331 | ɛ̱ | Latin Small Letter Open E + Combining Macron Below |
| 0254 + 0331 | ɔ̱ | Latin Small Letter Open O + Combining Macron Below |
| 0072 + 0303 | r̃ | Latin Small Letter R + Combining Tilde |
| 0289 + 0303 | ʉ̃ | Latin Small Letter U with Bar + Combining Tilde |

In each case below, the letter is followed by another letter (or two, in the case of two combining marks. (In each case shown, the letters were simply copied, then the font changed.)

Ariel Courier New

g̃g g̃g Latin Small Letter G + Combining Tilde

r̃r r̃r Latin Small Letter R + Combining Tilde

ɔ̱o ɔ̱o Latin Small Letter Open O + Combining Macron Below

ɛ̱e ɛ̱e Latin Small Letter Open E + Combining Macron Below

ɛ̱̈ee ɛ̱̈ee Latin Small Letter Open E + Combining Diaresis + Combining Macron

Below

ɨ̃i ɨ̃i Latin Small Letter I with Stroke + Combining Tilde

ʉ̃u ʉ̃u Latin Small Letter U with Bar + Combining Tilde

ẹ̀e ẹ̀e Latin Small Letter E with Dot Below + Combining Grave Accent

ẹ́e ẹ́e Latin Small Letter E with Dot Below + Combining Acute Accent

ọ̀o ọ̀o Latin Small Letter O with Dot Below + Combining Grave Accent

ọ́o ọ́o Latin Small Letter O with Dot Below + Combining Acute Accent

Findings:

* With each of these cases, error is a certainty. The ideal solution, of course, would be for the Unicode folks to create new pre-composed code points for these problem cases. But I suspect there is little chance of them doing so before our report is due. So we will have to figure out an alternate approach to recommend. The simplest would be to just remove these glyphs from our repertoire, ending updates to Unicode. But perhaps we can find a better way.