Proposal to Adjust Identifier Types for Certain Modifier Letters

Asmus Freytag

2021-07-23

Several scripts have modifier letters or spacing tone marks that are indistinguishable from Punctuation, including punctuation found directly in fully qualified domain names or punctuation that may be used to terminate or surround domain names.

While they are PVALID in IDNA2008, these should not have Identifier_Type "Recommended". Instead, it would be more appropriate to move them to "Inclusion". They may be appropriate (or harmless) in some identifier contexts, but are definitely problematic in domain names. The way "Inclusion" is derived in UTS#39 would require these characters to be added to Table 3a in UAX#31; or for the definition to be expanded to allow these as an additional set.

Independently, these should also be added to the UTS#39 data file "Intentional.txt" with these entries.

0027 ; A78C # (' ~ ') APOSTROPHE ~ LATIN SMALL LETTER SALTILLO

002C ; A4F9 # (, ~ ,) COMMA ~ LISU LETTER TONE NA PO

002E ; A4F8 # (. ~ .) FULL STOP ~ LISU LETTER TONE MYA TI

002E ; 1C79 # (. ~ .) FULL STOP ~ OL CHIKI GAAHLAA TTUDDAAG

003B ; A4FC # (; ~ ;) SEMICOLON ~ LISU LETTER TONE MYA NA

003C ; A4FD # (: ~ :) COLON ~ LISU LETTER TONE MYA JEU

007C ; 01C0 # (|~ |) VERTICAL BAR ~ MODIFIER LETTER CLICK

02BB ; 2018 # (' ~ ') MODIFIER LETTER TURNED COMMA ~ LEFT SINGLE QUOTATION MARK
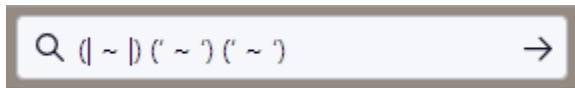
02BC ; 2019 # (' ~ ') MODIFIER LETTER APOSTROPHE ~ RIGHT SINGLE QUOTATION MARK

It should be noted that both comma and semicolon show greater variation in shapes depending on the font used and some fonts may render them slightly differently from the way a Lisu or Ol Chiki font would render the tone marks, but as this screenshot from a browser search bar shows, they are effectively indistinguishable in practice:

Their shapes in Lisu were clearly direct borrowings from standard (Latin) punctuation, which further supports their classification as "intentionally" the same appearance. Note further that Lisu, while using Western punctuation, does not use the period because of confusion with the tone mark.

Likewise, while the rendering of quotation marks and apostrophes can vary across fonts, we find that harmonized fonts do not make any distinction between the punctuation and the modifier letters (or the Saltillo), as shown in these screen shots.

As a result, despite being PVALID, these letters have been excluded from the DNS Root Zone for security reasons.