

Collation Adjustments for DUCET

Author: Richard Wordingham
Date: 21 May 2012

The UTC has just agreed (UTC 131-A27) to add the following contractions to DUCET:

0FB2 0F71 ; [.255A.0020.0002.0FB2] [.2570.0020.0002.0F71]
0FB3 0F71 ; [.255D.0020.0002.0FB3] [.2570.0020.0002.0F71]

I believe these may be expressed as a tailoring

& \u0fb2\u0f71 = \u0fb2\u0f71
& \u0fb3\u0f71 = \u0fb3\u0f71

I have tested the collation of a range of Tibetan vowels and found that these additions changed the collation orders of some perfectly valid Tibetan sequences:

0FBC 0F72 and 0FB2 0F71 0F72
0FBC 0F74 and 0FB2 0F71 0F74
0FB3 0F72 and 0FB3 0F71 0F72
0FB3 0F74 and 0FB3 0F71 0F74

Disappointingly, neither of the tests defined by CollationTest_NON_IGNOREABLE.txt and CollationTest_SHIFTED.txt detected these transpositions.

I therefore propose that another 4 tailorings be added to undo this accidental re-ordering and re-establish their previous ordering:

& \u0fb2\u0f71\u0f72 = \u0fb2\u0f71\u0f72
& \u0fb2\u0f71\u0f74 = \u0fb2\u0f71\u0f74
& \u0fb3\u0f71\u0f72 = \u0fb3\u0f71\u0f72
& \u0fb3\u0f71\u0f74 = \u0fb3\u0f71\u0f74

These should be applied before the two new tailorings.

The intention is that if no other changes were made to DUCET, these should add the following entries to DUCET:

0FB2 0F73 ; [.255A.0020.0002.0FB2] [.2572.0020.0002.0F73]
0FB2 0F71 0F72 ; [.255A.0020.0002.0FB2] [.2572.0020.0002.0F73]
0FB2 0F75 ; [.255A.0020.0002.0FB2] [.2576.0020.0002.0F75]
0FB2 0F71 0F74 ; [.255A.0020.0002.0FB2] [.2576.0020.0002.0F75]
0FB3 0F73 ; [.255D.0020.0002.0FB3] [.2572.0020.0002.0F73]
0FB3 0F71 0F72 ; [.255D.0020.0002.0FB3] [.2572.0020.0002.0F73]
0FB3 0F75 ; [.255D.0020.0002.0FB3] [.2576.0020.0002.0F75]
0FB3 0F71 0F74 ; [.255D.0020.0002.0FB3] [.2576.0020.0002.0F75]

In particular, any collation-based search for U+0F73 should continue to find it in in the sequence <U+0FB2, U+0F73>.