UCA DUCET: change tertiary weights for combining letters

Author: Markus Scherer  
Date: 2012-sep-12

Proposal

Change collation decomposition types (decomps.txt) for most combining letters (see Details section below) from <sort> to <super>/<sub>, and change their tertiary weights accordingly.

Rationale

In UCA 6.2, the following sort the same:

1. dz U+01F3 LATIN SMALL LETTER DZ
2. ⚷ U+02A3 LATIN SMALL LETTER DZ DIGRAPH
3. ͩ U+0369 COMBINING LATIN SMALL LETTER D + U+1DE6 COMBINING LATIN SMALL LETTER Z
4. ⅾ U+217E SMALL ROMAN NUMERAL FIVE HUNDRED + U+1DE6 COMBINING LATIN SMALL LETTER Z

Most of these are ok, but it seems strange that the dz ligature (line 1) sorts the same as a sequence beginning with U+0369 Combining Latin Small Letter d (line 3) which is a combining mark and looks more like a superscript.

All such combining letters are treated like “vanilla” compatibility variants, which makes them sort like other such variants. The proposal is to treat the combining letters like superscript or subscript variants, as appropriate.

Alternatives

It would be possible to do one of the following:

1. Not make any changes.
2. Use the DUCET reserved tertiary weight 0007.
3. Change these to have secondary differences rather than tertiary ones.

Details

The combining letters proposed to be changed have primary collation elements and sort like variants of regular letters.

Change the following collation decomposition types to <super>

0363;<sort>;0061 # COMBINING LATIN SMALL LETTER A => LATIN SMALL LETTER A
0364;<sort>;0065 # COMBINING LATIN SMALL LETTER E => LATIN SMALL LETTER E
0365;<sort>;0069 # COMBINING LATIN SMALL LETTER I => LATIN SMALL LETTER I
Note: It looks like this would be good for some Tai Tham characters as well, for example U+1A59 & U+1A5A. There are many combining marks in this script, and I do not have enough information at this time to make an informed proposal for Tai Tham.

Change the following collation decomposition type to <sub>

1DCA; <sort>; 0072 # COMBINING LATIN SMALL LETTER R BELOW => LATIN SMALL LETTER R