

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. đz 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
```

0366;<sort>;006F # COMBINING LATIN SMALL LETTER O => LATIN SMALL LETTER O
0367;<sort>;0075 # COMBINING LATIN SMALL LETTER U => LATIN SMALL LETTER U
0368;<sort>;0063 # COMBINING LATIN SMALL LETTER C => LATIN SMALL LETTER C
0369;<sort>;0064 # COMBINING LATIN SMALL LETTER D => LATIN SMALL LETTER D
036A;<sort>;0068 # COMBINING LATIN SMALL LETTER H => LATIN SMALL LETTER H
036B;<sort>;006D # COMBINING LATIN SMALL LETTER M => LATIN SMALL LETTER M
036C;<sort>;0072 # COMBINING LATIN SMALL LETTER R => LATIN SMALL LETTER R
036D;<sort>;0074 # COMBINING LATIN SMALL LETTER T => LATIN SMALL LETTER T
036E;<sort>;0076 # COMBINING LATIN SMALL LETTER V => LATIN SMALL LETTER V
036F;<sort>;0078 # COMBINING LATIN SMALL LETTER X => LATIN SMALL LETTER X

1DD2;<sort>;A76F # COMBINING US ABOVE => LATIN SMALL LETTER CON
1DD3;<sort>;0061 F8F0 # COMBINING LATIN SMALL LETTER FLATTENED OPEN A ABOVE => LATIN SMALL LETTER A + PSEUDO-COMBINING VARIANT MARK 1
1DD4;<sort>;0061 F8F0 0065 # COMBINING LATIN SMALL LETTER AE => LATIN SMALL LETTER A + PSEUDO-COMBINING VARIANT MARK 1 + LATIN SMALL LETTER E
1DD5;<sort>;0061 006F # COMBINING LATIN SMALL LETTER AO => LATIN SMALL LETTER A + LATIN SMALL LETTER O
1DD6;<sort>;0061 0076 # COMBINING LATIN SMALL LETTER AV => LATIN SMALL LETTER A + LATIN SMALL LETTER V
1DD7;<sort>;0063 0327 # COMBINING LATIN SMALL LETTER C CEDILLA => LATIN SMALL LETTER C + COMBINING CEDILLA
1DD8;<sort>;0064 F8F1 # COMBINING LATIN SMALL LETTER INSULAR D => LATIN SMALL LETTER D + PSEUDO-COMBINING VARIANT MARK 2
1DD9;<sort>;0064 F8F0 # COMBINING LATIN SMALL LETTER ETH => LATIN SMALL LETTER D + PSEUDO-COMBINING VARIANT MARK 1
1DDA;<sort>;0067 # COMBINING LATIN SMALL LETTER G => LATIN SMALL LETTER G
1ddb;<sort>;0262 # COMBINING LATIN LETTER SMALL CAPITAL G => LATIN LETTER SMALL CAPITAL G
1DDC;<sort>;006B # COMBINING LATIN SMALL LETTER K => LATIN SMALL LETTER K
1DDD;<sort>;006C # COMBINING LATIN SMALL LETTER L => LATIN SMALL LETTER L
1DDE;<sort>;029F # COMBINING LATIN LETTER SMALL CAPITAL L => LATIN LETTER SMALL CAPITAL L
1DDF;<sort>;1D0D # COMBINING LATIN LETTER SMALL CAPITAL M => LATIN LETTER SMALL CAPITAL M
1DE0;<sort>;006E # COMBINING LATIN SMALL LETTER N => LATIN SMALL LETTER N
1DE1;<sort>;0274 # COMBINING LATIN LETTER SMALL CAPITAL N => LATIN LETTER SMALL CAPITAL N
1DE2;<sort>;0280 # COMBINING LATIN LETTER SMALL CAPITAL R => LATIN LETTER SMALL CAPITAL R
1DE3;<sort>;A75B # COMBINING LATIN SMALL LETTER R ROTUNDA => LATIN SMALL LETTER R ROTUNDA
1DE4;<sort>;0073 # COMBINING LATIN SMALL LETTER S => LATIN SMALL LETTER S
1DE5;<sort>;0073 F8F1 # COMBINING LATIN SMALL LETTER LONG S => LATIN SMALL LETTER S + PSEUDO-COMBINING VARIANT MARK 2
1DE6;<sort>;007A # COMBINING LATIN SMALL LETTER Z => LATIN SMALL LETTER Z

2DE0;<sort>;0431 # COMBINING CYRILLIC LETTER BE => CYRILLIC SMALL LETTER BE
2DE1;<sort>;0432 # COMBINING CYRILLIC LETTER VE => CYRILLIC SMALL LETTER VE
2DE2;<sort>;0433 # COMBINING CYRILLIC LETTER GHE => CYRILLIC SMALL LETTER GHE
2DE3;<sort>;0434 # COMBINING CYRILLIC LETTER DE => CYRILLIC SMALL LETTER DE
2DE4;<sort>;0436 # COMBINING CYRILLIC LETTER ZHE => CYRILLIC SMALL LETTER ZHE
2DE5;<sort>;0437 # COMBINING CYRILLIC LETTER ZE => CYRILLIC SMALL LETTER ZE
2DE6;<sort>;043A # COMBINING CYRILLIC LETTER KA => CYRILLIC SMALL LETTER KA
2DE7;<sort>;043B # COMBINING CYRILLIC LETTER EL => CYRILLIC SMALL LETTER EL
2DE8;<sort>;043C # COMBINING CYRILLIC LETTER EM => CYRILLIC SMALL LETTER EM
2DE9;<sort>;043D # COMBINING CYRILLIC LETTER EN => CYRILLIC SMALL LETTER EN
2DEA;<sort>;043E # COMBINING CYRILLIC LETTER O => CYRILLIC SMALL LETTER O

2DEB;<sort>;043F # COMBINING CYRILLIC LETTER PE => CYRILLIC SMALL LETTER PE
 2DEC;<sort>;0440 # COMBINING CYRILLIC LETTER ER => CYRILLIC SMALL LETTER ER
 2DED;<sort>;0441 # COMBINING CYRILLIC LETTER ES => CYRILLIC SMALL LETTER ES
 2DEE;<sort>;0442 # COMBINING CYRILLIC LETTER TE => CYRILLIC SMALL LETTER TE
 2DEF;<sort>;0445 # COMBINING CYRILLIC LETTER HA => CYRILLIC SMALL LETTER HA
 2DF0;<sort>;0446 # COMBINING CYRILLIC LETTER TSE => CYRILLIC SMALL LETTER TSE
 2DF1;<sort>;0447 # COMBINING CYRILLIC LETTER CHE => CYRILLIC SMALL LETTER CHE
 2DF2;<sort>;0448 # COMBINING CYRILLIC LETTER SHA => CYRILLIC SMALL LETTER SHA
 2DF3;<sort>;0449 # COMBINING CYRILLIC LETTER SHCHA => CYRILLIC SMALL LETTER SHCHA
 2DF4;<sort>;0473 # COMBINING CYRILLIC LETTER FITA => CYRILLIC SMALL LETTER FITA
 2DF5;<sort>;0441 0442 # COMBINING CYRILLIC LETTER ES-TE => CYRILLIC SMALL LETTER ES + CYRILLIC SMALL LETTER TE
 2DF6;<sort>;0430 # COMBINING CYRILLIC LETTER A => CYRILLIC SMALL LETTER A
 2DF7;<sort>;0435 # COMBINING CYRILLIC LETTER IE => CYRILLIC SMALL LETTER IE
 2DF8;<sort>;A649 # COMBINING CYRILLIC LETTER DJERV => CYRILLIC SMALL LETTER DJERV
 2DF9;<sort>;A64B # COMBINING CYRILLIC LETTER MONOGRAPH UK => CYRILLIC SMALL LETTER MONOGRAPH UK
 2DFA;<sort>;0463 # COMBINING CYRILLIC LETTER YAT => CYRILLIC SMALL LETTER YAT
 2DFB;<sort>;044E # COMBINING CYRILLIC LETTER YU => CYRILLIC SMALL LETTER YU
 2DFC;<sort>;A657 # COMBINING CYRILLIC LETTER IOTIFIED A => CYRILLIC SMALL LETTER IOTIFIED A
 2DFD;<sort>;0467 # COMBINING CYRILLIC LETTER LITTLE YUS => CYRILLIC SMALL LETTER LITTLE YUS
 2DFE;<sort>;046B # COMBINING CYRILLIC LETTER BIG YUS => CYRILLIC SMALL LETTER BIG YUS
 2DFF;<sort>;046D # COMBINING CYRILLIC LETTER IOTIFIED BIG YUS => CYRILLIC SMALL LETTER IOTIFIED BIG YUS

A674;<sort>;0454 # COMBINING CYRILLIC LETTER UKRAINIAN IE => CYRILLIC SMALL LETTER UKRAINIAN IE
 A675;<sort>;0438 # COMBINING CYRILLIC LETTER I => CYRILLIC SMALL LETTER I
 A676;<sort>;0457 # COMBINING CYRILLIC LETTER YI => CYRILLIC SMALL LETTER YI
 A677;<sort>;0443 # COMBINING CYRILLIC LETTER U => CYRILLIC SMALL LETTER U
 A678;<sort>;044A # COMBINING CYRILLIC LETTER HARD SIGN => CYRILLIC SMALL LETTER HARD SIGN
 A679;<sort>;044B # COMBINING CYRILLIC LETTER YERU => CYRILLIC SMALL LETTER YERU
 A67A;<sort>;044C # COMBINING CYRILLIC LETTER SOFT SIGN => CYRILLIC SMALL LETTER SOFT SIGN
 A67B;<sort>;0461 # COMBINING CYRILLIC LETTER OMEGA => CYRILLIC SMALL LETTER OMEGA

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