

UTC/1999-016

June 2, 1999

Title: Changes between UnicodeData-2.1.9d3.txt and UnicodeData-2.1.9.txt

Source: Ken Whistler, Sybase

Action: For consideration and approval by UTC

UnicodeData-2.1.9d3.txt was the delta version of UnicodeData that rolled up all the fixes agreed to at the UTC #79 meeting in Palo Alto. Since then, a number of other changes were applied and went into the final delta (UnicodeData-2.1.9d9.txt) that was posted as the UnicodeData.txt file associated with the 2.1.9 update.

The UTC needs to retroactively approve all normative changes that went into that final delta. All such changes and their rationales are explained below. For completeness, non-normative changes to the data file are also listed here.

1. [NORMATIVE] Change the combining class of two Tibetan marks from 230 to 220. This was a clear error and oversight, since the marks in question display underneath the base.

```
< 0F35;TIBETAN MARK NGAS BZUNG NYI ZLA;Mn;230;ON;;;;;N;TIBETAN HONORIFIC UNDER RING;nge zung nyi da;;;
< 0F37;TIBETAN MARK NGAS BZUNG SGOR RTAGS;Mn;230;ON;;;;;N;TIBETAN UNDER RING;nge zung gor ta;;;
---
> 0F35;TIBETAN MARK NGAS BZUNG NYI ZLA;Mn;220;ON;;;;;N;TIBETAN HONORIFIC UNDER RING;nge zung nyi da;;;
> 0F37;TIBETAN MARK NGAS BZUNG SGOR RTAGS;Mn;220;ON;;;;;N;TIBETAN UNDER RING;nge zung gor ta;;;

```

2. [NORMATIVE] A complex series of changes to several Tibetan vowel signs to eliminate some stanky edge cases for normalization. The essence of this was to change the "a-chung" U+0F71 from a combining class of 0 to a combining class of 129:

```
< 0F71;TIBETAN VOWEL SIGN AA;Mn;0;ON;;;;;N;;;;;
---
> 0F71;TIBETAN VOWEL SIGN AA;Mn;129;ON;;;;;N;;;;;

```

and then to reorder the canonical decompositions for U+0F73 and U+0F81, so that the decompositions would come out in canonically reordered form:

```
< 0F73;TIBETAN VOWEL SIGN II;Mn;0;ON;0F72 0F71;;;;;N;;;;;
< 0F81;TIBETAN VOWEL SIGN REVERSED II;Mn;0;ON;0F80 0F71;;;;;N;;;;;
---
> 0F73;TIBETAN VOWEL SIGN II;Mn;0;ON;0F71 0F72;;;;;N;;;;;
> 0F81;TIBETAN VOWEL SIGN REVERSED II;Mn;0;ON;0F71 0F80;;;;;N;;;;;

```

and finally to give U+0F77 and U+0F79 new **compatibility** decompositions, so that they would no recompose over the a-chung during normalization:

```
< 0F77;TIBETAN VOWEL SIGN VOCALIC RR;Mn;0;ON;0F76 0F71;;;;;N;;;;;
< 0F79;TIBETAN VOWEL SIGN VOCALIC LL;Mn;0;ON;0F78 0F71;;;;;N;;;;;
---
> 0F77;TIBETAN VOWEL SIGN VOCALIC RR;Mn;0;ON;<compat> 0FB2 0F81;;;;;N;;;;;

```

> 0F79;TIBETAN VOWEL SIGN VOCALIC LL;Mn;0;ON;<compat> 0FB3 0F81;;;N;;;;;

3. [INFORMATIVE] Change the category of the two Greek numeral signs from Po to Sk. This was done so that U+0374 matched in category U+02B9, to which is canonically equivalent. U+0375 was changed also for parallelism.

< 0374;GREEK NUMERAL SIGN;Po;0;L;02B9;;;N;GREEK UPPER NUMERAL SIGN;Dexia keraia;;;< 0375;GREEK LOWER NUMERAL SIGN;Po;0;L;;;;N;Aristeri keraia;;;

> 0374;GREEK NUMERAL SIGN;Sk;0;ON;02B9;;;N;GREEK UPPER NUMERAL SIGN;Dexia keraia;;> 0375;GREEK LOWER NUMERAL SIGN;Sk;0;ON;;;;N;Aristeri keraia;;;

4. [INFORMATIVE] The canonical decomposition mappings for a number of precomposed Latin characters were altered, so that the recursive decomposition automatically results in a canonically reordered combining character sequence. These are tone-marked Vietnamese letter combinations which had mappings that made linguistic sense, but that did not follow the claimed invariant of the UnicodeData file, namely that full decompositions are guaranteed to be in canonically reordered form.

< 1EAC;LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND DOT BELOW;Lu;0;L;00C2 0323;;;N;;;1EAD;< 1EAD;LATIN SMALL LETTER A WITH CIRCUMFLEX AND DOT BELOW;Ll;0;L;00E2 0323;;;N;;;1EAC;;1EAC< 1EB6;LATIN CAPITAL LETTER A WITH BREVE AND DOT BELOW;Lu;0;L;0102 0323;;;N;;;1EB7;< 1EB7;LATIN SMALL LETTER A WITH BREVE AND DOT BELOW;Ll;0;L;0103 0323;;;N;;;1EB6;;1EB6< 1EC6;LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND DOT BELOW;Lu;0;L;00CA 0323;;;N;;;1EC7;< 1EC7;LATIN SMALL LETTER E WITH CIRCUMFLEX AND DOT BELOW;Ll;0;L;00EA 0323;;;N;;;1EC6;;1EC6< 1ED8;LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND DOT BELOW;Lu;0;L;00D4 0323;;;N;;;1ED9;< 1ED9;LATIN SMALL LETTER O WITH CIRCUMFLEX AND DOT BELOW;Ll;0;L;00F4 0323;;;N;;;1ED8;;1ED8< 1E1C;LATIN CAPITAL LETTER E WITH CEDILLA AND BREVE;Lu;0;L;0114 0327;;;N;;;1E1D;< 1E1D;LATIN SMALL LETTER E WITH CEDILLA AND BREVE;Ll;0;L;0115 0327;;;N;;;1E1C;;1E1C

> 1EAC;LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND DOT BELOW;Lu;0;L;1EA0 0302;;;N;;;1EAD;> 1EAD;LATIN SMALL LETTER A WITH CIRCUMFLEX AND DOT BELOW;Ll;0;L;1EA1 0302;;;N;;;1EAC;;1EAC> 1EB6;LATIN CAPITAL LETTER A WITH BREVE AND DOT BELOW;Lu;0;L;1EA0 0306;;;N;;;1EB7;> 1EB7;LATIN SMALL LETTER A WITH BREVE AND DOT BELOW;Ll;0;L;1EA1 0306;;;N;;;1EB6;;1EB6> 1EC6;LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND DOT BELOW;Lu;0;L;1EB8 0302;;;N;;;1EC7;> 1EC7;LATIN SMALL LETTER E WITH CIRCUMFLEX AND DOT BELOW;Ll;0;L;1EB9 0302;;;N;;;1EC6;;1EC6> 1ED8;LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND DOT BELOW;Lu;0;L;1ECC 0302;;;N;;;1ED9;> 1ED9;LATIN SMALL LETTER O WITH CIRCUMFLEX AND DOT BELOW;Ll;0;L;1ECD 0302;;;N;;;1ED8;;1ED8> 1E1C;LATIN CAPITAL LETTER E WITH CEDILLA AND BREVE;Lu;0;L;0045 0327 0306;;;N;;;1E1D;> 1E1D;LATIN SMALL LETTER E WITH CEDILLA AND BREVE;Ll;0;L;0065 0327 0306;;;N;;;1E1C;;1E1C

5. [INFORMATIVE] A similar change in decomposition mappings for a number of compatibility Arabic ligatures involving SHADDA. The full decomposition does not change -- only the decomposition mapping in the table, so that the recursive decomposition is already in canonically reordered form.

< FC5E;ARABIC LIGATURE SHADDA WITH DAMMATAN ISOLATED FORM;Lo;0;R;<isolated> 0020 0651 064C;;;N;;;;< FC5F;ARABIC LIGATURE SHADDA WITH KASRATAN ISOLATED FORM;Lo;0;R;<isolated> 0020 0651 064D;;;N;;;;< FC60;ARABIC LIGATURE SHADDA WITH FATHA ISOLATED FORM;Lo;0;R;<isolated> 0020 0651 064E;;;N;;;;< FC61;ARABIC LIGATURE SHADDA WITH DAMMA ISOLATED FORM;Lo;0;R;<isolated> 0020 0651 064F;;;N;;;;< FC62;ARABIC LIGATURE SHADDA WITH KASRA ISOLATED FORM;Lo;0;R;<isolated> 0020 0651 0650;;;N;;;;< FCF2;ARABIC LIGATURE SHADDA WITH FATHA MEDIAL FORM;Lo;0;R;<medial> 0640 0651 064E;;;N;;;;;

```

< FCF3;ARABIC LIGATURE SHADDA WITH DAMMA MEDIAL FORM;Lo;0;R;<medial> 0640 0651 064F;;;N;;;;
< FCF4;ARABIC LIGATURE SHADDA WITH KASRA MEDIAL FORM;Lo;0;R;<medial> 0640 0651 0650;;;N;;;;
---
> FC5E;ARABIC LIGATURE SHADDA WITH DAMMATAN ISOLATED FORM;Lo;0;R;<isolated> 0020 064C 0651;;;N;;;;
> FC5F;ARABIC LIGATURE SHADDA WITH KASRATAN ISOLATED FORM;Lo;0;R;<isolated> 0020 064D 0651;;;N;;;;
> FC60;ARABIC LIGATURE SHADDA WITH FATHA ISOLATED FORM;Lo;0;R;<isolated> 0020 064E 0651;;;N;;;;
> FC61;ARABIC LIGATURE SHADDA WITH DAMMA ISOLATED FORM;Lo;0;R;<isolated> 0020 064F 0651;;;N;;;;
> FC62;ARABIC LIGATURE SHADDA WITH KASRA ISOLATED FORM;Lo;0;R;<isolated> 0020 0650 0651;;;N;;;;
> FCF2;ARABIC LIGATURE SHADDA WITH FATHA MEDIAL FORM;Lo;0;R;<medial> 0640 064E 0651;;;N;;;;
> FCF3;ARABIC LIGATURE SHADDA WITH DAMMA MEDIAL FORM;Lo;0;R;<medial> 0640 064F 0651;;;N;;;;
> FCF4;ARABIC LIGATURE SHADDA WITH KASRA MEDIAL FORM;Lo;0;R;<medial> 0640 0650 0651;;;N;;;;

```