Proposal to add a Stability Policy: No more decompositions to combining marks

Author: Markus Scherer
Date: 2012-mar-15

I propose the addition of the following Stability Policy:

> ***Applicable Version: Unicode 2.1+***
>
> **No further characters will be encoded whose canonical decomposition starts with a non-zero combining mark.**
>
> Unicode has 7 such characters (U+0340, U+0341, U+0343, U+0344, U+0F73, U+0F75, U+0F81).
>
> Such characters add significant complexity for processing processing of canonically equivalent text.

Note that such characters are not very useful since they get normalized away even in NFC.

Sample issues:

- Enumerating the set of strings that are canonically equivalent to an input string.
- Collation contractions with non-leading combining marks of a decomposition mapping.

Details: http://unicode.org/cldr/utility/list-unicodeset.jsp?a=[[:nfd_qc=no:]%26[:^lccc=0:]]&g=age

Age=1.1

Combining Diacritical Marks — *Vietnamese tone marks*

U+0340 ( ) COMBINING GRAVE TONE MARK
U+0341 ( ) COMBINING ACUTE TONE MARK

Combining Diacritical Marks — *Additions for Greek*

U+0343 ( ) COMBINING GREEK KORONIS
U+0344 ( ) COMBINING GREEK DIALYTIKA TONOS

Age=2.0

Tibetan — *Dependent vowel sign*

U+0F73 ( ) TIBETAN VOWEL SIGN II
U+0F75 ( ) TIBETAN VOWEL SIGN UU
U+0F81 ( ) TIBETAN VOWEL SIGN REVERSED II