

A Request to revise the current model for encoding Bengali Khanda-Ta

In the Unicode Standard, version 4, an attempt to address the problem of encoding the Bengali letter Khanda-Ta has been made. Section 9.2 states,

“The Unicode Standard provides an explicit way to encode a half-letter form.”...
...“the sequence U+09A4, U+09CD, U+200D is displayed as a khanda ta glyph ঁ”

From this statement it can be seen that the standard considers Khanda-Ta to be similar to a half form.

Khanda-Ta actually behaves in a manner very different to a half-letter. An example of this fact is that vowel signs that are normally reordered around letters must not reorder around Khanda-Ta.

To understand the problem with the current specification I would like you to examine the following examples. Lets us assume that a Unicode input stream contains the following sequence:

U+09A4 U+09CD U+09B2 U+09C7
ত + ঁ + ল + ঐ
(Ta + Virama + La +VowelSignE)

1. In a comprehensive font this sequence may displayed as a Ta-La conjunct with the vowel reordered to the left :

ঐ
VowelSignE + Ta_La.conjunct

2. Because the Ta-La conjunct is quite rare, it quite likely that a font may not contain this conjunct. Sometimes in cases like this, a fall back convention of using a half+full form conjunct is used.

ঐ
VowelSignE + HalfTa + La (ঐ+ত+ল)

3. Alternatively the Khanda-Ta is can be used:

ঐ
Khanda-Ta+ VowelSignE + La

In this case, the VowelSign must not reorder around the Ta but around the La instead.

It is not considered practical that a display engine should be expected to know what conjuncts are available in any given font. If a font defaults to displaying a Khanda-Ta due to lack of a full conjunct glyph (and with the current Unicode specification that is the only half form option available) how can we expect the display logic to know about this so that it can reorder a vowel signs in an appropriate way. Seeing as a rendering device cannot literally see what is displayed, there *is no* way.

So far no one has been able to suggest a work-around that's addresses this issue. (e.g the suggestion of encoding KhandaTa as Ta+VIRAMA+ZWJ+ZWNJ found in document [L2/03-209](#) is only a partial work around)

In addition, the specification of the 'Bengali Half-Ta' as 'Khanda-Ta' has made it impossible to specify a half-sized version of Ta in Unicode. Seeing as Bengali has the ability to encode half versions of all its other letters, it is not seen why 'Ta' should be an exception.

The solution to this problem was solved long ago by the implementers of ISCII. But so far Unicode has not been willing to accept this answer. I will repeat the solution here :-)

Khanda-Ta is directly equivalent to Ta+Virama. At some point in history Khanda-Ta became the practised way of writing Ta+Virama . (It was the cool way to write the letter Ta when it was the last letter of a word, and inherent 'a' was usually suppressed at the end of words)

ISCII implementations encoded it as such i.e. Khanta-Ta is the default form to be displayed when ever Ta+Virama is encountered. When the Ta+Virama is followed by another consonant it is always displayed as either half-(size) Ta or a full conjunct depending on the font. (But never as Khanda-Ta). Thus solving the problem.

In fact before the Unicode standard decided to state that Khanda-Ta was to be encoded as a half letter, there were working Unicode implementations satisfactorily using the ISCII model of Khanda-Ta. They worked as follows:

The 'Halant' form of Ta was Khanda-Ta (triggered by Ta+Virama+(space) or Ta+Virama+ZWNJ)

The Half form of Ta was a Half sized Ta) (triggered by Ta+Virama+ZWJ). (In some fonts HalfTa looked like Ta+Virama).

(In the unusual event that a Ta with a visible virama need be displayed it can be encoded as TA ZWNJ VIRAMA)

Andy.