ISO/IEC JTC1/SC2/WG2 N3606
2009-03-30
L2/09-110

Doc Type: Working Group Document
Title: Proposal to Advance the Renaming of Arabic Pedagogical Symbols
Source: Anshuman Pandey (pandey@umich.edu)
Status: Individual Contribution
Action: For consideration by ISO/IEC JTC1/SC2/WG2
Date: 2009-03-30

Sixteen Arabic Pedagogical Symbols (U+FBB2..U+FBC1) were accepted for encoding in October 2008. The names of the characters as per Amendment 7 (pages 30–31 of N3557):[1]

```
FBB2 ARABIC SINGLE NUQTA ABOVE
FBB3 ARABIC SINGLE NUQTA BELOW
FBB4 ARABIC DOUBLE NUQTA ABOVE
FBB5 ARABIC DOUBLE NUQTA BELOW
FBB6 ARABIC TRIPLE NUQTA ABOVE
FBB7 ARABIC TRIPLE NUQTA BELOW
FBB8 ARABIC TRIPLE INVERTED NUQTA ABOVE
FBB9 ARABIC TRIPLE INVERTED NUQTA BELOW
FBBA ARABIC QUADRUPLE NUQTA ABOVE
FBBB ARABIC QUADRUPLE NUQTA BELOW
FBBC ARABIC DOUBLE DANDA BELOW
FBBD ARABIC DOUBLE NUQTA VERTICAL ABOVE
FBBE ARABIC DOUBLE NUQTA VERTICAL BELOW
FBBF ARABIC SINGLE CIRCLE BELOW
FBC0 ARABIC TOTA ABOVE
FBC1 ARABIC TOTA BELOW
```

In February 2009, Roozbeh Pournader recommended revising the above character names in his "Proposal for consistent naming of Arabic Pedagogical Symbols" (N3575).[2] He writes that "the character names for the newly proposed characters are not consistent with names of existing Arabic characters in ISO/IEC 10646. This proposal suggests consistent names for the characters. The consistency would also help making the Unicode Standard easier to use for people unfamiliar with Pakistani terminology presently used for those accepted characters." Pournader suggests the following names for the characters:

```
FBB2 ARABIC SYMBOL DOT ABOVE
FBB3 ARABIC SYMBOL DOT BELOW
FBB4 ARABIC SYMBOL TWO DOTS ABOVE
FBB5 ARABIC SYMBOL TWO DOTS BELOW
FBB6 ARABIC SYMBOL THREE DOTS ABOVE
FBB7 ARABIC SYMBOL THREE DOTS BELOW
FBB8 ARABIC SYMBOL THREE DOTS POINTING DOWNWARDS ABOVE
FBB9 ARABIC SYMBOL THREE DOTS POINTING DOWNWARDS BELOW
FBBA ARABIC SYMBOL FOUR DOTS ABOVE
FBBB ARABIC SYMBOL FOUR DOTS BELOW
FBBC ARABIC SYMBOL TWO DANDAS BELOW
FBBD ARABIC SYMBOL TWO DOTS VERTICALLY ABOVE
FBBE ARABIC SYMBOL TWO DOTS VERTICALLY BELOW
FBBF ARABIC SYMBOL RING
FBC0 ARABIC SYMBOL SMALL TAH ABOVE
FBC1 ARABIC SYMBOL SMALL TAH BELOW
```

[1] http://std.dkuug.dk/jtc1/sc2/wg2/docs/n3557.pdf
[2] http://std.dkuug.dk/jtc1/sc2/wg2/docs/n3575.pdf

Pournader's suggested changes for the names of the Arabic Pedagogical Symbols reflect the existing naming convention for Arabic characters.

However, it should be noted that there is nothing wrong with the use of the term NUQTA in the names of these characters. NUQTA is simply a transliteration of the original Arabic نقطة *nuqṭā*. This *nuqṭā* is the basis for NUKTA (Devanagari नुक़्ता *nuqtā* > *nuktā*), which is the name for the ़ character encoded in the UCS as SIGN NUKTA in nearly every north Indic script. Thus, there is a precedent for using NUQTA in character names, but as NUKTA, and not for Arabic. The preference in Arabic has been to use DOT and DOTS. This precedent is reflected in Pournader's suggested names.

I am in agreement with Pournader's suggested changes, but recommend the following:

1. The name for FBBC should be changed to ARABIC SYMBOL DOUBLE VERTICAL BAR BELOW. The use of the term DANDA in Pournader's name ARABIC SYMBOL TWO DANDAS BELOW evokes the ubiquitous Indic punctuation mark. The name is also problematic because a more 'precise' name would be *ARABIC SYMBOL DOUBLE DANDA BELOW. There is no tradition of using the character or the term DANDA in Arabic. In order to avoid the semantic confusion of characters and names across scripts, a generic name should be chosen for FBBC. The name ARABIC SYMBOL DOUBLE VERTICAL BAR BELOW is generic, free from associations with other characters from other scripts, and fully descriptive.

2. FBC0 and FBC1 should be annotated to indicate that term TOTA is used for these characters, as shown in N3557. This is an orthographic term used in Urdu for the SMALL TAH component of various letters and is not appropriate for describing a character that may be used to represent other languages, or for a representation form of a character that is already represented in the UCS: ط U+0637 ARABIC LETTER TAH. However, طوطا *ṭoṭā* is not solely "Pakistani terminology"; the term is used for the SMALL TAH symbol in India and in other locations where Urdu is used. Nor is it specifically an Urdu term; it is derived from Arabic.

   It is easy to imagine a pedagogical case in Urdu orthography or Nastaliq calligraphy that describes the composition of ٹ U+0679 ARABIC LETTER TTEH:

   > The letter ٹ *ṭe* is produced by writing ب (U+066E ARABIC LETTER DOTLESS BEH) the shape of *be* with ◌ٹ *totā* (U+FBC0 ARABIC SYMBOL SMALL TAH ABOVE).

   Since FBC0 and FBC1 are pedagogical symbols, they should be annotated to indicate that they are "known as 'tota' in Urdu". Such annotations will help to facilitate recognition of the characters for those who will use these symbols for producing pedagogical materials.

I strongly recommend that the character names for the Arabic Pedagogical Symbols be changed as proposed by Pourander with modifications to his suggestions as made in my recommendations above.