

To: UTC

From: Roozbeh Pournader (forwarded by Debbie Anderson, SEI, UC Berkeley)

Subject: Allocating Arabic Extended-C in SMP and Arabic code point changes

Date: 25 August 2021

Background:

More Arabic proposals are expected. Because the BMP is getting filled up, I think we need to reserve some space for some characters like potential *hehs* for Sindhi and Kurdish in the BMP, and put the historical and technical characters in the SMP. Lorna Evans concurs and the Script Ad Hoc was agreeable to the suggestion.

Proposal:

I propose allocating 10EC0..10EFF to Arabic Extended-C, and move some of post-14 Arabic characters there. A good rule of thumb for choosing BMP for a character would be modern non-technical usage.

Below is my suggestion for codepoints to be moved to the SMP:

Characters already approved by UTC

- ARABIC SMALL LOW WORD SAKTA 0895 to 10EFD
- ARABIC SMALL LOW WORD QASR 0896 to 10EFE
- ARABIC SMALL LOW WORD MADDA 0897 to 10EFF

Codepoints for pending characters:

- Western Cham mark to 10EF0
- FEH WITH THREE DOTS POINTING DOWNWARDS ABOVE at 10EC0
- NOON WITH TWO DOTS VERTICALLY ABOVE at 10EC1

This leaves 7 empty slots in the Arabic Extended-B block for important letters and combining marks. We will also have 28 slots in Arabic Presentation Forms-A and -B blocks for things that don't interact, like pedagogical symbols, honorifics, and even punctuation.