

To UTC
From Kamal Mansour
Date 2002-04-26

L2/02-161

Progress Report on L2/02-21 (Proposal to amend Arabic repertoire)

Presented at UTC 90 (Feb 2002), L2/02-21 proposes the addition to the Unicode repertoire of certain diacritics that, in general, meet at least one of the following conditions:

1. Repeatedly used in combination with various base letters.
2. Needed for the formation of yet unencoded, but necessary, letter-diacritic combinations.

During the discussion of the proposal, there was concern that the addition of such characters may disturb the normalization conventions already in place for Arabic script. To determine in which way to push forward with the discussion, the UTC asked me to assess the potential number of Arabic characters that have yet to be encoded. It is expected that unencoded, Arabic-script characters will most likely belong to minority languages. Of course, making such an assessment can be difficult because written material of such minority languages does not typically gain broad circulation.

Several leads pointed me to two printed references that contain a systematic listing of the Arabic-character repertoires of many languages. The first one [Maji1984], a published thesis by Majidi that covers a broad range of languages, confirms that the current Unicode Arabic-script repertoire does indeed satisfy the needs of a great many languages. However, I did manage to find 3 Kashmiri letters that could only be represented in Unicode as of the UTC meeting of February 2002.

The second source [Chta1992] is a monograph by Chtatou that deals with African languages written in Arabic script. The author surveys and critiques a set of proposals for transcribing numerous African languages. In conclusion, he regularized and rationalized proposed characters by searching for phonetic similarities among the various languages, as well as by using existing letter combinations derived from Farsi, Urdu, Pashto, and Jawi usage. The examples demonstrate a significant number of cases that use existing diacritics in new combinations for consonants, in addition to the introduction of new diacritics for vowels. Although these efforts at transcribing African languages do demonstrate some interesting trends, I do not consider the information sufficient for making a general assessment. For the next few months, I plan on searching for additional material in order to be able to make a conclusive recommendation to the UTC.

[Chta1992] Mohamed Chtatou. *Using Arabic Script in Writing the Languages of the Peoples of Muslim Africa*. Publications of the Institute of African Studies. Rabat, Morocco. 1992.

[Maji1984] Mohammad-Reza Majidi. *Das arabisch-persische Alphabet in den Sprachen der Welt: eine graphemisch-phonemische Untersuchung*. H. Buske Verlag. Hamburg, Germany. 1984.