Requirements and Process for Changing Script Status for Identifier Use

Asmus Freytag

January 7, 2021

UTS#39 defines Identifier type Recommended as characters in "widespread common everyday use". Formally, the definition is based on membership of the character in a Recommended script in UAX#31 (with some exceptions). Recommended scripts are therefore in "widespread common everyday use", while other scripts with less active modern use might be classed as Limited_Use.

These characterizations are not permanent; they are intended to track actual use of a given script, including any significant changes in usage. The definition of Recommended script is used as input to other specifications outside of the Unicode Standard, such as the Label Generation Rules for the DNS Root Zone (see "Root Zone LGR" under https://icann.org/idn for details).

Because of such dependencies, it is advisable to use a very deliberate process when adjusting the status of a script in UAX#31 (and therefore the Identifier_Type of its member characters). Such a process must first and foremost establish whether the usage for a Limited_Use script has changed sufficiently so that it fits the requirements of being in "widespread common everyday use".

In principle, a small user community alone does not disqualify, but no matter the size of that community, there are a number of factors that weigh in favor or against the proposition that the script is in "widespread everyday common use".

A lot of data is readily available for individual languages and their user community, but also the degree to which the language is in active use (for example: Ethnologue) and being actively transmitted to the next generation of speakers (see Expanded Graded Intergenerational Disruption Scale or EGIDS). With some care, like factoring literacy data and the effect of competing scripts, such data can be a useful proxy for some of the information that may not yet have been compiled on the script level.

Scripts that are used for languages with an EGIDS level of 4 or smaller, and particularly if no alternate scripts serve the same language communities in day-to-day contexts could be assumed to be in "widespread everyday common use". Additional supportive evidence for that supposition would be active support in social media, Wikipedia, commercial, administrative and governmental websites, etc. in the given script. If, on the contrary, the script is primarily used in the preservation of cultural heritage with day-to-day activities conducted in other (recommended) scripts, then that would argue against making a change in classification at this time.

The questions a proposal to reclassify a script should address:

1. What is the community served by the script proposed for reclassification? (Region, countries)

2. What are the principal languages? (Or does the user community consist only of scholars/not-native speakers?)

3. What is the number of native speakers? (for the principal language(s) using the script)

4. How many of them are literate? (In which scripts?)

5. What other scripts are these languages written in? (List all)

6. Are any of the languages 'recognized' for official use?

7. Is the script 'recognized' for official use?

8. Is there evidence of official business conducted in the languages? In what scripts? (List all!) (Provide examples)

9. Is there educational activity for any of the languages? In what scripts? (List all!) (Provide examples/citations)

10. Is there evidence of significant levels of day-to-day business carried out using one or more of these languages in the proposed script? (Provide examples)

11. Is there evidence of publishing activity in these languages in the proposed script? (Provide examples/citations)

12. Is there a Wikipedia in one or more of the languages? Is it written in the proposed script? (Provide link)

13. Is there a country/organization that would be likely to apply for a top level or domain name in the proposed script?

14. For domains that allow registrations in the proposed script today, have any domain names been registered in the script?

15. Is the script universally supported for display / input on desktop platforms? (Which ones are missing?)

16. Is the script universally supported for display / input on mobile platforms? (Which ones are missing?)

17. Does any of the support require installing aftermarket or supplemental fonts/keyboards?

18. Is there any divergence in rendering across platforms?

19. Is the encoding model for the script fully stable?

20. Is there a mail host and/or social media platform which support user names in this script? (List all known)

21. Is there any hard data that makes the case that user communities have been/continue to shift to using the script?