L2/26-219

Supporting the needs of communities with newly invented scripts

Jan Kučera

2026-10-21

Summary: This document proposes establishing an ad-hoc group to evaluate the feasibility of Unicode or SEW providing script packages for applications keen on supporting emerging writing systems.

Introduction

The Script Encoding Working Group (SEW) is receiving an increasing number of proposals for new scripts invented rather recently, including scripts younger than 10 or even 5 years. Usually, these proposals fall short on two of the three main criteria for encoding (stability and evidence of usage), while putting considerable emphasis on the need for digital interchange. Moreover, some of the communities are heavily phone-oriented and have no interest in producing printed publications. The existence of printed publications is not a requirement for encoding, but it is the easiest way to demonstrate that the objectives of the criteria have been met.

The ultimate reason for the criteria for encoding is to ensure that the encoded characters will be used, forever. Encoding new characters is not free, it comes with considerable costs. On the Unicode side, preparation of data, updating algorithms, text for Core Specification, charts for publication, as well as coordination with the ISO and production of their standard. On the implementers' side, it requires the development of fonts, keyboards, tests, sometimes updates to the implementations of algorithms and shaping engines, deploying to all devices, committing to bug fixing and supporting the script, forever. On the users' side, there are increased storage, memory and processing requirements, forever.

We are seeing growing frustration of communities in the amount of time it takes to get their script on their devices. Not only it can take a few years to encode a script in Unicode, but it can then take several years for the implementers to support the new Unicode version, design, implement and ship all the required infrastructure, including securing the necessary budget. Furthermore, some older devices might not be eligible for updates anymore.

We are also seeing politics and identity, both within communities and outside of them, playing an increasing role in having particular scripts encoded or not encoded. It is becoming more necessary as well as difficult for the SEW to verify the claims made in encoding proposals, ensure it represents everyone it claims to represent, and does not cause more harm than good. For a small volunteer organization, this is simply not possible with the current resources available.

Drawbacks of premature encoding

The SEW and UTC are under pressure to speed up the encoding process and lower the bar for encoding. However, this can lead to undesirable consequences for both Unicode and the communities it is trying to serve.

On the Unicode side, the implementers have entrusted the consortium to do its due diligence and ensure that the costs going into encoding new scripts do not go in vain. If Unicode were to be over-eager and encodes characters that aren't used, this causes implementers to incur the costs above with no final value, or destabilizes users. If this happens frequently, implementers could decide to stop taking in new versions of Unicode, which jeopardizes the very pipeline which communities are trying to get their scripts encoded into.

On the user communities' side, it needs to be understood that once a script is encoded, several core characteristics are frozen, forever. That includes the encoding model, the character names and some of their other properties, and to some extent, the representative glyphs. When the community is still trying to disseminate a new script and is figuring out what will stick with users, what will not, and what will need adjustments, it is too late to be doing these changes after the script is encoded. This is not negotiable; stability of the encoding is the sole purpose of standardization and strictly enforced policies are in place to uphold this core value that Unicode provides to implementers as well as to its users.

Alternative avenues

The SEW members care about communities around the world and would like to better support them in their efforts and help them make a stronger case for encoding. After discussions within the group, the SEI and several proposal authors, it was suggested that a system similar to software library packages could be developed to provide support of emerging scripts in individual applications.

It is not the goal to provide a system-wide support for new scripts, competing or conflicting with scripts encoded in Unicode. It has been observed that the largest demand for script encoding in the communities is to interchange short messages amongst users using the same application, and that system-wide support is prohibitively slow to get to the end users.

It is expected that the packages could be loaded into an application at runtime and would include at least the following in one form or another:

- Font file under Open Font License
- Character properties needed for shaping by the Universal Shaping Engine
- LDML keyboard layout file

The SEW recognizes that presenting text to users using third-party fonts and/or properties comes with considerable security risks. However, Unicode is uniquely positioned to ensure that the packages published by Unicode meet the security requirements of implementers and guarantee that by means of digital signatures or similar.

Such a system of packages would allow communities to quickly iterate on their script and stabilize it faster based on real data.

Having Unicode or SEW provide the system would allow the SEW to have reliable evidence on how much a script is being used by the community, and gain support for the packages from Unicode members and other implementers.

Feasibility study

It is proposed an ad-hoc group to be established under the UTC to explore the feasibility of providing script packages for applications. The mandate of the group would be as follows:

- Define the goals and non-goals of script packages
- Explore possible ways of implementation, their requirements for implementers and limitations for users
- Seek feedback from vendors on which requirements could be realistically met
- Seek feedback from communities on which limitations would not jeopardize the usefulness of the solution
- Consider whether Unicode can provide the resources and infrastructure for distributing script packages

The ad-hoc group will conclude with a report to the UTC on the considered options and whether the idea is worth pursuing.