



AL-ALAC-ST-0724-02-00-EN
ORIGINAL: English
DATE: 26 August 2024
STATUS: Ratified

AT-LARGE ADVISORY COMMITTEE
ALAC Statement on the Preliminary Issue Report on Latin Script Diacritics

Preamble

On 18 July 2024 the Public Comment proceeding opened for the [Preliminary Issue Report on Latin Script Diacritics](#). An At-Large workspace was created in preparation for this Public Comment submission. The At-Large Consolidated Policy Working Group (CPWG) reviewed the Preliminary Issue Report and decided it would be in the interest of end users to develop and submit an ALAC Public Comment Statement.

Preamble.....1
Ratification Record.....2
ALAC and At-Large Comments.....2

Ratification Record

On 18 July 2024, the Public Comment proceeding opened for the Preliminary Issue Report on Latin Script Diacritics. On 24 July 2024, Bill Jouris, Satish Babu, and Patrick Kiiza (penholders) volunteered to review the Preliminary Issue Report. On 14 August 2024, the penholders discussed the comments for the ALAC statement during the CPWG call. The At-Large Consolidated Policy Working Group (CPWG) decided it would be in the interest of end users to develop and submit an ALAC statement for this Public Comment proceeding. Penholders volunteered to draft the initial ALAC statement.

On 14 August 2024, the initial draft statement was shared with the CPWG for review and input. On 21 August 2024, the At-Large Public Comment Statement was finalized. The ALAC Chair, Jonathan Zuck, requested that the Public Comment Statement be ratified by the ALAC before submission to the ICANN Public Comment feature.

On 26 August 2024, staff confirmed the online vote resulted in the ALAC endorsing the statement with 15 out of 15 votes in favor. 0 votes against, and 0 abstentions. Please note 100% of ALAC members participated in the poll. The ALAC members who participated in the poll are (alphabetical order by first name): Aziz Hilali, Bill Jouris, Bukola Oronti, Claire Craig, Eduardo Diaz, Joanna Kulesza, Jonathan Zuck, Justine Chew, Lilian Ivette De Luque, Marcelo Rodriguez, Pari Esfandiari, Raihanath Gbadamassi, Satish Babu, Shah Zahidur Rahman, and Tommi Karttaavi. You may view the results here:
<https://tally.icann.org/cgi/results?e=3989a645acb>.

ALAC and At-Large Comments

The ALAC and the At-Large community appreciate the opportunity to comment on the Preliminary Issue Report on Latin Script Diacritics.

The matter of Diacritics in the Latin Script is of significant importance to end-users. Many prominent language and script communities, especially in Europe and the Americas, have pointed out the lack of support in the DNS to letters that are considered equivalent in general use.

We understand that the Latin Generation Panel (GP) took the position of blocking most of the code points that are visually indistinguishable, on account of an abundance of caution arising out of the fact that the GP had to consider 200+ scripts, and security and stability of the root was the most important consideration.

The matter of “.québec” has been raised numerous times by end-users, and more formally communicated to ICANN by Point Quebec recently. We would like to point out that there are possibly many other parties who would have liked to register diacritics-based versions of their names in addition to their ASCII versions, who have simply abandoned their plans because there is no way to register those names as variants. An enabling policy, where diacritics can be

managed as equivalents of ASCII codepoints, will significantly benefit these script communities, applicants and end-users.

The ALAC and the At-Large community therefore strongly support the strategy outlined in the Preliminary Issues Report, which is for a single-issue Policy Development Process (PDP) to create consensus policy that would enable the use of diacritics. In addition, we suggest that the scope of the PDP include the consideration for the timely introduction of any such consensus policy (if any) that is unconstrained by the timing of the New gTLD Program application windows.

Considering that the MSR-5 contains 312 code points for Latin¹, it may be prudent to limit the approved diacritic-equivalents to the minimum, perhaps based on an analysis of use. This will ensure that the number of variants in Latin script is minimized, which is in line with the Principle of Conservatism, a fundamental principle identified in the Phase 1 report of EPDP on IDNs.

Additional Comments

1. It would have been impossible for the Latin Generation Panel to identify pairs which were “usually considered equivalent,” given the number of languages involved and the fact that the members of the panel were not fluent in the vast majority of them. (Note that the examples given in Section 3.3.1 do not include *any* African languages which use diacritics. This reflects the severity of the problem.) Instead, the Panel only identified pairs which a reader would find it difficult or impossible to distinguish at a normal font size. That is, visual variants. As a result, no letter would be a variant of that letter plus diacritic – the diacritic, no matter what it was, would make the pair visually distinct.
2. The complexity is that we are batching all "Latin Script" for over 200 languages across the whole world. Each language has very different rules on the use of diacritics and some even have special rules in relation to the same diacritic used in different words. We would ask that the PDP address the challenges of dealing with that complexity.
3. The expected time for a single-issue PDP would likely run past the scheduled date for the next round. An approach that will deal with this is needed.
4. Finally, we draw your attention to the table on top of Page 8, which provides equivalent words with and without diacritics. Such letter pairs as ‘e’ and ‘é’ or ‘a’ and ‘à’ are indeed considered as equivalents in general use. However, after consulting our Spanish-speaking colleagues, it appears that ‘n’ and ‘ñ’ are not equivalents (ie., they are not considered as variants of each other). Both of them are distinct letters of the Spanish language. For instance, ‘unas’ and ‘uñas’ are both valid words in Spanish but they have completely different meanings. Therefore, ‘manana’ would not be considered a valid word in general use in Spanish.

¹ The top three scripts with the largest numbers of code points in MSR-5 are Han, Hangul and Latin (Source: P 41 of MSR-5 revision 24 June 2021)