

Preliminary Issue Report on Latin Script Diacritics

Update to CPWG, 31 July 2024

Bill Jouris, Kiiza Patrick, Satish Babu

Background

- Québec, the largest province of Canada, is predominantly Francophone. The official language of the province is French
- Point Quebec (PQ) had applied for the Geoname “.quebec” in 2012, and the name was delegated in 2014
- Since a majority of local users used the accented form of the gTLD, PQ had been wanting the delegation of “.québec” as well as an alternative to “.quebec”
- There are only two ways to achieve this, and both options are infeasible at this time

Option 1: Apply as a variant

- The Phase 1 Recommendations of the EPDP on IDNs provide the simplest way to address this issue: apply as a variant, and then manage the two labels as a variant set
- The Latin GP provides guidance on what variants could be supported. Their position was perhaps influenced by the following factors:
 - The large number of languages covered (200+) in the GP
 - The absence of native speakers in most of them in the GP team
 - The need to be conservative about variants
 - Security and stability were high priority
- Due to these factors, there were very few allocatable variants identified (mostly obscure letter pairs such as “i” (0131) and ‘i’ (0069))

Option 1 (contd)

- The “Variant Principles Matrix” (p. 37) of the Latin LGR 7 states that “Visual variants (generally acceptable alternate glyphs)” are blocked on account of Security
- Therefore, none of the letter pairs usually considered as equivalents (such as ‘e’ and ‘é’) were available in the LGR as variants
- This makes it impossible for “.québec” and “.quebec” to be variants

Option 2: “.québec” as a new application

- The second option is for PQ to apply for “.québec” as a separate application and then manage both together (of course, PQ would have to pay the base application fee again)
- However, this is likely to pose issues at the String Similarity Review step in the application process
- There is a very high chance that the applied-for name, “.québec” would be considered confusingly similar to an existing delegated name “.quebec”, and would be rejected
- Even in the case of new applications (such as “.puglia” and “.pùglia”), from the same applicant, both would end up in a contention set because of their similarity
- Therefore, this approach doesn’t solve the problem either

ALAC's efforts

- Point Quebec brought this issue to ALAC's attention, as there was no way to address it in the current system
- Some of the ALAC members met with representatives of PQ and ICANN Staff in meetings during ICANN76 and ICAN77
- On 22 June 2023, the ALAC Chair wrote to the GNSO Council Chair pointing out the issue and asking for resolution
- On 16 May 2024, the GNSO Council resolved to request for an Issue Report and directed staff to create the Report

Importance to end-users

- While it was PQ that advocated the need for a practical way to address the problem, the issue is actually more widespread. ALAC's letter quotes three examples: “.cafe” and “.café”; “.hermes” and “.hermès”; and “.hotels” and “.hôtels”
- In many cases, applicants have dropped the diacritic version of names simply because there was no way to do it prior to IDNs being available
- Many business, institutions and geographic entities, especially in countries in Europe, North America and Latin America, would benefit from a new policy
- Providing a proper process to achieve this will help end-users and applications in creating a more “natural” Domain Name System

The run-up to the Issue Report

- Prior to asking for an Issue Report, GNSO Council had asked ICANN Org to identify any mechanism (short of a PDP) to mitigate the issue
- ICANN Org proposed to leverage “...non-adopted recommendations related to string similarity since in essence, a solution for this issue is likely an exception process”. In other words, they proposed an alternative pathway during String Similarity Review
- However, the Council “...was not comfortable with this solution and instead requested an Issue Report”
- The Council was particularly interested in a variants-based approach, since the Phase 1 report of the EPDP on IDNs provided a clear pathway

The Issue Report

- A background to diacritics and their use
- The Latin GP's decisions and the aftermath
- The String Similarity Review process
- The Problem Statement and scope
- Applicability of EPDP Phase 1 and 2 Recommendations
- Consideration of applicability of ccTLD's "exceptional" solution
- Potential issues to be considered in a PDP:
 - Single issue PDP
 - Possible impact on Human Rights

Questions (3.2.4)

- Under what circumstances should a base ASCII gTLD and the Latin script diacritic version of the gTLD be simultaneously delegated, if any?
 - If such circumstances exist, what measures should be put into place in order to mitigate the potential for end-user confusion?
- If a solution is needed to this issue, are any of the elements of the ccTLD solution transferable?
- If a solution is needed to this issue, are any of the elements from either Phase 1 or Phase 2 of the EPDP on IDNs relevant, or warrant discussion specific to Latin script diacritics?

Outcomes (3.2.3)

- The limited circumstances in which a base ASCII gTLD and the Latin script diacritic version of the gTLD can be simultaneously delegated.
- The rules governing instances where the ASCII/Latin script diacritic are simultaneously delegated.

Challenges

- The Latin GP's work touches upon 212 scripts/languages that are used by about 70% of the literate world
- The GP has taken a very conservative approach to the question of variants
- However, opening up the LGR too widely may also be quite destabilizing and has a non-trivial risk of end-user confusion
- Given the existence consensus policy on variants, a variants-based approach may be best suited to solve the issue of Diacritics
- The main challenge appears to be in managing boundaries of how many code points are to be variant-enabled

Small Group Recommendations

- The Issues Report proposes a new, single-topic PDP to resolve the issue of diacritics
- The report also hints at a preference for a variants-based approach, although the final decision on this is to be taken by the PDP
- In our opinion, ALAC should, through a statement,
 - Highlight the end-user aspects of diacritics including the scale and impact
 - Support the course of action proposed in the Issue Report
 - Highlight any factual errors in the report
- The report does not mention how PDP was chosen as the vehicle (as against an EPDP). Does CPWG think the latter is more appropriate? Is there a time criticality in this case?

Thank you