Preliminary Issue Report on Latin Script Diacritics
27-Aug-2024
Comment of the ICANN Business Constituency (BC)

Overview

The omission of Latin diacritics in the DNS has been a workaround inherited from historical ASCII limitations and does not signify an accurate or preferred representation of words by the involved linguistic communities. Adopting rules that reflect the advancements in Universal Acceptance in the past decade acknowledges the priority the ICANN community has consensually agreed upon to increase linguistic diversity on the Internet.

The Latin script presents unique challenges that have remained unaddressed since the previous round of new GTLDs and preempting the repetition of conflicts around the Latin script is a task that the GNSO needs to undertake with priority in order for these rules to integrate with the process of the upcoming round of new GTLDs and become part of it as soon as possible.

Similar issues have been addressed in the context of IDN ccTLDs, such as the case of .ευ and .eu. ICANN developed an exception procedure that allowed for the simultaneous management of these visually similar but distinct labels by the same registry operator. Similar solutions need to be found for GTLDs, as the underlying technology and motives are the same.

Expectations

This comment advocates for the ability for applicants to propose and manage TLD pairs in ASCII and IDN/diacritic versions, allowing them to be controlled as unit by the same operator, under specific conditions set by the community that address potential user confusion.

This should take place within the context of a tightly focused PDP with limited scope and the clear goal of addressing the ASCII and IDN overlaps that take place in some instances where diacritics are used in the Latin alphabet. No other asks/issues should be bundled with the PDP, and it should follow a straight path aiming for clear community consensus.

We find the outline provided by the Issue Report to be sufficient and to set adequate expectations around this PDP and would like to encourage the swift establishment of the processes required for this matte to advance.

Our expectation is that if both the ASCII and IDN strings are applied-for strings, they will not end up placed into a string contention set with each other. Instead, the unique requirements of this use case need to be acknowledged and understood within proper context.

Moreover, it will be beneficial if these two strings co-exist and be treated as a parent (ASCII) and child (variant) relationship, wherein domain registration in the parent be automatically replicated in the child as a derivative option with the same registrant. There will be no need to create the child registration utilizing "EPP-create" transaction hence the registry / registrar therefore should not charge additional fee to the registrant.

Practical examples

In Brazilian Portuguese, there are a variety of cases that exemplify the two different ways in which the issue at hand can affect GTLDs, of which we will present two below:

Interchangeable: ".saopaulo" and ".sãopaulo" can be both used to represent the Brazilian city of São Paulo and this is easily understood by people who make use of Latin-based languages, without the potential for confusion. This, in practice, is the same issue presented in the leading case that affects this PDP, which is that of ".quebec" and ".québec".

Non-interchangeable: "*.maca*" (stretcher)" and "*.maçã*" (apple). In this case, the ASCII version, "maca" corresponds directly to the word utilized for stretchers, making into a TLD suitable for medicine or sports. Meanwhile, it is possible to make use of the same base characters, but with diacritics added, to create a completely different word that stands for the fruit apple. In this case, there is no exchangeability and the coexistence of both TLDs as separate and distinct makes sense.

This comment was drafted by Mark Datysgeld and Vivek Goyal.

It was approved in accord with our charter.