Dear IP,

The Latin GP appreciates the additional input received on August 29, 2018 titled “Diacritics below a security risk?”.

In its feedback, IP makes a case for potential security risks when certain diacritics below (e.g. dot and line below) are used in a domain name. The risk is not always apparent, but it reveals itself when the diacritic below is obscured by an underline, which is the typical formatting feature of hyperlinks. The IP asserts “**Of all diacritics, diacritics below can be difficult to distinguish or be prone to clipping”.**

Security risks deserve a place in our analysis, so Latin GP is committed to explicitly discuss this matter, resolve whether such risks constitute an actual security risk to the Root Zone and decide a solution vis-à-vis the LGR.

The GP is however unclear how many risks should be dealt with at the level of LGR vs. post-LGR review processes, such as String Similarity Review, DNS Stability Review processes and is therefore requesting additional guidance from IP regarding earlier communiquées which seem to be at odds with the most recent one.

Thus far, the Latin GP has been aiming to deal with the “straight forward, non-subjectives cases” of visual similarity (i.e. homoglyphs).

Specifically, the Latin GP notes these statements from the August 29 email:

*“It can be argued users have no working understanding of typography and would not reliably interpret small gaps or bulges in the underline as being related to an unfamiliar code point”*.

*“The IP would like to encourage the Latin GP […] to explicitly examine [the diacritics below] example and other cases like it, where code points can become indistinguishable in common usage scenarios for IDNs”*.

In this context, the Latin GP wants to confirm that prior guidance from IP (below), which we find consistent with the LGR Procedure, is not at odds with the August 29 email.

                LGR Procedure:

“Finally, in investigating the possible variant relations, Generation Panels should ignore cases where the relation is based exclusively on aspects of visual similarity.”

October 18, 2017: Principles for repertoire and variants – Feedback from IP

*“In the context of the Root Zone, the Procedure is quite clear in that it considers simple similarity of appearance to be outside the scope of the Root Zone LGR.”*... *“Having the Root Zone exhibit fundamentally different design decisions with respect to variants than those found on the second level would have to be justified by strong arguments based on factors special to the Root Zone.”*

March 22, 2017: Latin GP Proposal: IP Feedback

*“The kinds of variants to be defined in the Root Zone LGR are limited to homoglyphs, which are characters essentially identical appearance by design, instead of merely similar appearance.”*

It may be helpful in clarifying matters if the IP could address a few actual examples for us.

We are all clear that

o  and ỗ     006F and 1ED7

are sufficiently different to not even be confusable.

Where in between does the threshold for variants lie:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Case no | Code point 1 | Glyph 1 | Glyph 2 | Code point 2 | Variants or "merely visually similar"  or confusable or something else |
|  | 0131 | ı | ɩ | 0269 |  |
|  | 00E1 | ă | ǎ | 01CE |  |
|  | 006A | j | į | 012F |  |
|  | 00FC | ü | ű | 0171 |  |
|  | 006C | m | rn | 0072 +  006D |  |
|  | 006B | k | ƙ | 0199 |  |
|  | 0070 | p | þ | 00FE |  |
|  | 0064 | d | ɗ | 0257 |  |
|  | 00F5 | õ | ō | 014D |  |
|  | 0069 | i | ı | 0131 |  |
|  | 1E35 | ḵ | k | 006B |  |
|  | 0137 | ķ | k | 006B |  |
|  | 014D | ō | o | 006F |  |
| 14. | 1E37 | ḷ | l | 006C |  |
| 15. | 010B | ċ | c | 0063 |  |
| 16. | 006C | m | м | 043C |  |
| 17. | 014B | ŋ | η | 03B7 |  |
| 18. | 0062 + 006C | bl | ы | 044B |  |

FYI, Case no 10 in the previous table, an experiment with a couple dozen native speakers of English found them unable to spot occurrences of the latter.  Even when told that there was a variation on the letter I somewhere in the text.