Internationalized Domain Names Expedited Policy Development Process

Rec 2.6, String Similarity Review Hybrid Model, D1b

IDN-EPDP Team Meeting #54 | 13 October 2022

Agenda

1. Roll Call and SOI Updates (2 mins)
2. Welcome and Chair Updates (5 min)
3. Review of Revised Project Plan (15 mins)
4. Review Rationale Language for Recommendation 2.6 (20 mins)
5. Continued Discussion of String Similarity Review Hybrid Model (45 mins)
6. Charter Question D1b (time permitting)
7. AOB (3 mins)

Review of Project Change Request

Option 1: One Phased Approach (Updated)

Timeline Consideration for One Phased Approach (Initial Report with Two Parts)

- Conservative estimate
- Additional time for reviewing operational input from ICANN org (10 weeks)
- Extended time for completing data collection to facilitate Phase 2 deliberation (8 months, Nov 2022 - Jul 2023)
- Extended time for Phase 2 charter question deliberation (3 weeks per question)
- Extended time for finalizing all final recommendations to produce one Final Report (18 weeks)

Key Milestone Date Projections for One Phased Approach

- Initial Report Part 1 Public Comment: 21 April 2023
- Initial Report Part 2 Public Comment: 3 January 2025
- Final Report submission to Council: **29 August 2025**

Option 2: Two Phased Approach

Timeline Consideration for Two-Phased Approach (Phase 1 on top-level & Phase 2 on second-level)

- Additional time required to develop two Initial Reports, process two public comments, and develop two Final Reports
  - Time period to finalize each phase of final recommendations after Public Comment may be shorter (12 weeks)
- Same timeline extension applied: ICANN org input review, data collection effort, second-level deliberations

Key Milestone Date Projections for Two Phased Approach

- Phase 1 Initial Report Public Comment: 21 April 2023
- Phase 1 Final Report submission to Council: **3 November 2023**
- Phase 2 Initial Report Public Comment: 25 April 2025
- Phase 2 Final Report submission to Council: **7 November 2025**

Caveat: *Since the EPDP Team is not expected to start Phase 2 deliberation until November 2023, this projected completion date is subject to change, based on observation of the team's progress during Phase 2 deliberations*

Continued Discussion of String Similarity Hybrid Model

Small Team Recommendation: Hybrid Model

Summary: *The small group recommends the* ***hybrid model****, which is a* ***mixed-level approach between level 2 and level 3***

Goal: *Mitigate any possibility of confusing similarity between one IDN TLD and another IDN TLD or any of its valid variant(s), vice versa*

In practice, the string similarity review must be modified to compare:

- **An applied-for primary IDN gTLD and <u>all of its allocatable variant label(s)</u>**

Against:

- **Existing TLDs and <u>all of their allocatable and blocked variant labels</u>;**
- **Strings requested as IDN ccTLDs and <u>all of their allocatable and blocked variant labels</u>;**
- **Other applied-for gTLDs in the same round and <u>all of their allocatable and blocked variant labels</u>;**
- **Reserved Names; and**
- **Any other two-character ASCII strings and <u>all of their allocatable and blocked variant labels</u> (*if the applied-for primary IDN gTLD is a two-character string*)**

In addition, compare:

- **<u>All of the blocked variant label(s)</u> of an applied-for primary IDN gTLD**

Against:

- **Existing TLDs and <u>all of their allocatable variant labels</u>**

Note: *Blocked variants of one IDN TLD should NOT be compared against blocked variants of another IDN TLD*

ALAC Position

The ALAC Team **supports the proposed Hybrid Model** for String Similarity Review on the basis that:

1. There is a need to adopt a cautious and conservative approach for the String Similarity review process.
2. The proposed Hybrid Model reflects a conservative approach that can help mitigate the two types of failure modes (denial of service and misconnection), and thereby promotes good user experience; yet does not call for excessive computational and evaluative complexity (by not comparing blocked variants of one TLD with the blocked variants of another).

We are, however, **open to considering refinements that would help reduce complexity in computation/ evaluation** without diminishing the mitigatory goal of the Hybrid Model.

GAC Position

In general we are **supportive, like other constituencies, of the Hybrid Model for String Similarity**.

While we take, perhaps, a less conservative position than some, looking for maximum opportunities for variant use, we appreciate the need to mitigate failures; not least through denial or service or misconnection. We note that neither of these failure modes would be beneficial to users.

We look forward to further work in the wider group on taking this forward.

Finally, we would like to thank the Small Group for all their work on this.

NCSG Position

1. We think that the **proposal of the small group is a practical one and backed by fact**.
2. The findings clearly shows that there is a real possibility to miss confusable strings if the applied for TLD was not compared against (a) its blocked variants (b) existing TLDs and their allocatable and blocked variants (c) other applied for TLDs and their allocatable and blocked variants.
3. A blocked variant of the request for TLD cannot be allocated, therefore we agree with the group's recommendation that it should not be used to compare against blocked variants of other TLDs. Only applied for and allocatable strings should be used.

RySG Position

The RySG agrees that the Small Group's recommendation is conservative, and that this is a desirable goal, along with the inclusion and stability principles, for DNS labels considered for the root.

The RySG also acknowledges that the Small Team did not consider the operational and procedural implications of the recommended Hybrid Model for the string similarity review process.

In this regard, the RySG would like to submit the following observations which may be considered in the IDN EPDP deliberations to finalize the policy recommendation and/or implementation guidelines:

1. The procedure to develop the RZ-LGR was not designed to identify visually similar labels as a primary objective; It may capture some cases but not all of them. For this reason, we recommend that this EPDP should be careful to not construe that the variant set produced by the RZ-LGR is a "complete" set of all visually similar labels; furthermore, if the intent is to capture and incorporate possible confusable labels of a gTLD, **the string similarity process should not rely solely on the output of the RZ-LGR (for visually confusable labels), but it may use it as an input to determine the basis for comparison**.
2. In general, **a conservative approach to processing candidate gTLDs for the root is a good approach**. That said, incorporating ineligible labels (blocked variants) into the string similarity process is unprecedented, so **careful consideration should be given to the implementation to achieve consistency and predictability of outcomes**.
3. As the Small Group noted, there are several reasons to mark a variant as "blocked". Some blocked variant labels are not well-formed labels e.g., mix-script labels, which do not conform to protocol or generally acceptable guidelines for DNS labels, including IDNs. Therefore, as long as mix-script labels are prohibited from gTLD labels (with the notable exception of Japanese), the RySG suggests that the **anticipated procedure should identify and remove mix-script labels from further consideration to avoid unnecessary work**.

IPC Views (Not Official Position)

With respect to the IPC view, we have not yet been able as a group to discuss this completely.

IPC is concerned that the **Hybrid Model needs to have an exception process**. For example, if there is a Brand that has a name which happens to be a non-allocatable variant of another string that has been delegated, then we believe that new Brand should be able to overcome any potential string similarity review of string confusion objection.

we believe that the **risk of "misconnection" in such a case is extremely low** in such a case.

Any model that does not allow for exceptions in our view would be unacceptable to the IPC.

Continued Discussion of D1b: Existing gTLD Registries Applying for Variants

Reflection on Strawman Process Flow

Origin of Strawman Process Flow

D1b: What should be the process by which an existing registry operator could apply for a variant for its existing gTLD?

B4: What should an application process look like in terms of timing and sequence for an existing and future Registry Operator with respect to applying or activating their allocatable variant TLD labels?

Observations

1. Understand which elements in the New gTLD Program will be impacted by variant implementation

An application for an IDN variant label needs to go through the same stages/steps in the New gTLD Program, just like a regular gTLD application

2. Consider how such elements will need to be modified to accommodate variant gTLDs

Around half of the elements in the New gTLD Program will require specific consideration/modification, in accordance with the recommendations proposed by the EPDP Team, to accommodate variant gTLD applications

3. Analyze the level of efforts of evaluating variant applications and the associated cost/fees

Only 44 existing gTLDs (35 Chinese gTLDs and 9 Arabic gTLDs) have allocatable variants and their registry operators may potentially for them.

While most Chinese gTLD registry operators and two Arabic gTLD ROs who responded to the survey indicated interest in applying for variants, it will likely be expensive and impractical to develop a standalone round just to accommodate these registries, given the observations related to items 1-2 above. Existing ROs applying for variants during the application rounds will help address the cost recovery principle.

Similarly, it will likely be expensive and impractical to allow activation of variant gTLD labels between application rounds.