

Consideration on reduction of allocatable labels (v0.2)

Background

CGP, JGP, and KGP are collaborating to create Root zone LGR for CJK. CJK GPs' collaboration is not limited to among us, but also involving IP and ICANN. Because Han characters have several variants in certain language(s), the integrated LGR (LGR-2, c.f. terminology) may generate tens of variant labels. Inflation of Root zone by variant labels are not desired, therefore some mechanisms for reducing allocatable labels must be considered.

Problem statement

Many of Han characters in LGR-2 have 2 or more variants. If an applied-for label has 5 characters and each character has 2 variants, in extreme situations, theoretically it might generate 32 ($=2^5$) variant labels including applied-for label¹. According to the proposed CJK integration algorithm, all the variants defined by other GPs are imported and make their variant type to “allocatable” except for the cases where it is not defined in original repertoire. Therefore, for example, if all of them are allocatable, and as requested to delegate most of them by applicant, the Root zone size easily inflates. This situation is not desired, so reduction mechanism of allocatable variant labels in Root zone LGR is highly recommended. In addition, labels in Root zone should be predictable from end users' perspective. This requires that set of variant labels in Root zone must be consist from combination of variants used in daily life, and keep small as much as possible. The “allocatable” reduction mechanism also must take this into account.

Followings are some possible reduction mechanisms:

- (1) ICANN/IP prepares additional process for applicant or evaluation-after-application panel to select reasonable numbers of variant labels from generated “allocatable” variant labels.
- (2) Each GP struggles to reduce variants in LGR-1, for instance by reducing variant characters.

¹ It is unclear whether “applied-for label” is always allocatable.

- (3) After the first CJK integration process (first generation), each GP reproduce next generation LGR-1 with passive variant definitions imported from first generation LGR-2 and some proper variant subtypes for them.

Considerations for each possible mechanism

For (1), this mechanism requires human interaction which takes extra time and resources. Root zone LGR is intended to provide automatic method as much as possible, so there will be a contradiction. We will need further discussion to this mechanism. Note that this human interaction process will keep “delegated” variant labels in predictable range.

For (2), CGP had already employed their best effort to reduce variants from 19000+ to 12000+. JGP defined no variants. KGP are going to define very small set of variants. Therefore, it seems unfeasible or unrealistic to do this.

For (3), CGP had already defined simp/trad/both variant subtypes and prohibits mixture of simp/trad subtypes, which means the maximum number of allocatable labels is three. JGP investigated probability of numbers of generated variant labels and actual usage of variant labels from existing “Japanese.JP” domain names. And JGP tentatively concluded that definition of proper variant subtype is impossible as there seems to be no general common and predictable rule for applied-for-string regarding variants². We will need KGP’s feedback.

Considering the different variant disposition in different language environment, one suggestion is that, unless there is a clear number restriction algorithm under a given language tag, all of the allocatable labels should go to the evaluation-after-application phase, the originally applied label go delegated, the others should be studied carefully case by case. The number limitation on delegated labels, could be reached through human interaction process, taking full account of its semantic meaning, instead of a number limitation on allocatable labels.

² Statistics will be provided during the meeting.