

CJK Integration Algorithm (v0.4)

Precondition

- Each of CJK Generation Panels (GPs) generates LGR for each language TLD before integration.
 - CJK GPs pick up ideographic variants for CJK from domain name usage perspective.
 - CJK GPs don't elaborate ideographic variants for CJK from linguistic perspective.
- CJK GPs agree on the mechanism (steps) to integrate and extract each language (script) LGR.

Step 1: Each CJK GP generates its own LGR (hereinafter, LGR-alpha)

- LGR-alpha generation process is left to each CJK GP.
- LGR-alpha format must follow XML schema for LGR.
<<https://datatracker.ietf.org/doc/draft-ietf-lager-specification/>>
- Each code point must have individual <char> element (i.e. don't use <range> element).
- Each <char> element in LGR-alpha must have reflexive mapping as <var> element (i.e. each code point must have explicit variant type/subtype).
- WLE of Each LGR-alpha must not conflict (Conflict of WLE must be coordinated and solved at this step).

Step 2: CJK GPs collectively generate a merged table of each LGR-alpha (hereinafter, LGR-M)

1. Extract every <char> element tagged "sc:Hani" from each LGR-alpha.
 2. For each extracted <char> element, check the existence of another <char> element with the same code point ("cp" value), and if exists, merge them into one element. At this time, "type" attribute of <var> element must be removed.
 3. After the check was finished, record every merged <char> elements to LGR-M.
- Repertoire of LGR-M is the union of all sc:Hani in each CJK LGR-alpha.

- 36 ● Variants of each <char> elements in LGR-M is the union of all variants defined
- 37 for the code point in each CJK LGR-alpha.
- 38 ● LGR-M does not have following information.
- 39 ➤ Language tag.
- 40 ➤ Variant type/subtype attribute of <var> elements.
- 41 ➤ WLE (Whole Label Evaluation rules).

42

43 Step 3: Each CJK GP extract its original repertoire with integrated variants from
44 LGR-M.

45

- 46 1. For each <char> element in its LGR-alpha (hereinafter, $\{\text{char}::\text{LGR-alpha}\}$),
- 47 extract <char> element of the same code point (“cp” value) from LGR-M
- 48 (hereinafter, $\{\text{char}::\text{LGR-M}\}$).
- 49 1.1. For each <var> element in $\{\text{char}::\text{LGR-M}\}$ (hereinafter, $\{\text{var}::\text{LGR-M}\}$),
- 50 compare with <var> elements in $\{\text{char}::\text{LGR-alpha}\}$ (hereinafter,
- 51 $\{\text{var}::\text{LGR-alpha}\}$).
- 52 1.1.1. If $\{\text{var}::\text{LGR-M}\}$ has the same code point (“cp” value) with one of
- 53 $\{\text{var}::\text{LGR-alpha}\}$, then copy “type” attribute of corresponding
- 54 $\{\text{var}::\text{LGR-alpha}\}$ to $\{\text{var}::\text{LGR-M}\}$.
- 55 1.1.2. Otherwise, if $\{\text{var}::\text{LGR-M}\}$ has the same code point (“cp” value)
- 56 with any of <char> elements in LGR-alpha, set “type” attribute
- 57 with value “review” to $\{\text{var}::\text{LGR-M}\}$.
- 58 1.1.3. Otherwise, set “type” attribute with value “blocked” to $\{\text{var}::\text{LGR-}$
- 59 $\text{M}\}$, and record the code point of $\{\text{var}::\text{LGR-M}\}$ to Out-of-
- 60 Repertoire list (hereinafter, $\{\text{OoR-list}\}$).
- 61 1.2. Record $\{\text{char}::\text{LGR-M}\}$ to Integrated-Repertoire list (hereinafter, $\{\text{IR-list}\}$).
- 62 1.3. If $\{\text{IR-list}\}$ includes “review” type, execute either (a) return to Step 1 and
- 63 redefine appropriate type/subtype to corresponding <var> elements, or (b)
- 64 replace all “review” by “blocked”.

65

66 Step 4: Each CJK GP add “Out of Repertoire” code points for symmetry.

67

- 68 1. For each code point in $\{\text{OoR-list}\}$ (hereinafter, $\{\text{cp}::\text{OoR-list}\}$), extract <char>
- 69 element of the same code point (“cp” value) from LGR-M (hereinafter,
- 70 $\{\text{char}::\text{LGR-M}\}$).
- 71 1.1. For each <var> element in $\{\text{char}::\text{LGR-M}\}$ (hereinafter, $\{\text{var}::\text{LGR-M}\}$),

- 72 compare $\{\text{cp}::\text{OoR-list}\}$ and code point (“cp” value) of $\{\text{var}::\text{LGR-M}\}$.
73 1.2. If the two code points are the same, add “type” attribute with value “out-of-
74 repertoire-var” to $\{\text{var}::\text{LGR-M}\}$.
75 1.3. Otherwise, add “type” attribute with value “blocked” to $\{\text{var}::\text{LGR-M}\}$.
76 1.4. Record $\{\text{char}::\text{LGR-M}\}$ to $\{\text{IR-list}\}$.

77

78 **Step 5: Each CJK GP merge WLE in LGR-alpha into one.**

79

- 80 1. Each GP extract <rules> element from each LGR-alpha and merge them into one
81 WLE (generate integrated <rules> element, hereinafter, $\{\text{rules}::\text{LGR-M}\}$).
- 82 2. Each GP add following rule to $\{\text{rules}::\text{LGR-M}\}$ for handling “out-of-repertoire-
83 var” variant type.
84 <action disp=”invalid” any-variant=”out-of-repertoire-var” />

85

86 **Step 6: Each CJK GP generates integrated LGR (hereinafter, LGR-beta).**

87

- 88 1. Each GP extract preambles from its LGR-alpha (hereinafter, $\{\text{preamble}::\text{LGR-}$
89 alpha}).
 - 90 2. Each GP extract all <char> elements with “tag” value other than “sc:Hani” and
91 record to IR-list.
 - 92 3. Each GP merge $\{\text{preamble}::\text{LGR-alpha}\}$, IR-list and $\{\text{rules}::\text{LGR-M}\}$ into LGR-
93 beta.
- 94
- 95 ● In other words, this step replaces body of <data> element and <rules> element of
96 LGR-alpha to IR-list and $\{\text{rules}::\text{LGR-M}\}$ respectively.