

Variant Principles

Sarmad Hussain

Center for Language Engineering

Al-Khawarizmi Institute of Computer Science

Lahore, Pakistan

1

- If any of the principles apply for any joining context (initial, medial, final or isolated) for any of the writing styles in common use, it is applicable to the characters in all joining contexts – this is because contexts are not stable across all operating systems (i.e. two joined characters may seem broken on a mobile device)

2

- Two letter code points are variants if their shapes are the ***same*** (homographs) in at least on joining context

3

- Two letter code points are variants if their shapes are ***similar*** (not same) but are perceived as same characters in contemporary use by a language community, even if other language communities may consider them as separate characters – following from the conservative principles of the root and the root being a common resource for all communities

4

- These principles are also applicable if one or both of the variant candidates are a combination of code points (letter and combining mark(s))

5

- All code points in a variant set are variants of each other. Variant sets are formed by transitively combining variants, i.e. if A is a variant of B and B is a variant of C, then {A, B, C} forms a variant set, even if A and C may not qualify as variants based on other principles.

6

- There is no primary or preferred code point within a variant set. All variants are equal and in a symmetrical relationship with others within the variant set. This means that if {A, B, C} is a variant set, none of A, B or C get a preferential treatment in making a label and the choice will be arbitrary based on the applicant for the label.

Some Issues to Clarify

- Should visually similar code points be variants?
- Should visually distinct code points be variants?
- Should transitivity be followed selectively or comprehensively for Arabic script?
- How do we limit variant sets?
 - Code point level
 - Whole label rule
- How do we decide between allocatable and blockable variants?
- Should combining marks be allowed as variants?

Thank You!