

IRGN2206 version 1

Title: Proposal for Annotating Easily Confused Ideographs

Author: Henry Chan

Type: Individual Contribution to IRG #48 and UTC

Date: 2017/06/15

Proposal

A lot of easily confused ideographs are currently encoded in CJK Unified Ideographs and Extensions. It is proposed that rare ideographs with very similar shape should be marked up with an “informative note” to clarify the semantic meaning of such character, and a “cross reference” to the easily confused character.

Annotation is employed in nearly every other block in ISO/IEC10646, except for CJK Unified Ideographs and its extensions. Annotating Source Code Separated characters is not proposed because that is covered in the Annex of ISO/IEC10646.

Scope

- Only characters very similar in appearance would be annotated.
- Only characters where the semantic meaning should be distinguished should be annotated. For example, U+81F4 致 and U+26936 致 which are similar in shape an identical in meaning, would not be annotated.
- Only the rare character(s) would be annotated. (Refer to Group 3 and 4 in Examples)

Examples

Group 1

1.1 Annotate 𡗗 U+2125A with:

- archaic form of 5C01 封

→ 37B7 峯

1.2 Annotate 𡗗 U+37B7

- phonetic of 狸, 徯, 漉...

→ 2125A 𡗗

Group 2

2.1 Annotate 周 (U+20C6C) with

- archaic form of 5468 周

→ 2055B 周

2.2 Annotate 周 (U+2055B) with

- archaic form of 5BB3 害

→ 20C6C 周

Group 3

3.1 Annotate 凡 (U+2007D) with

- archaic form of 4E38 丸

→ 51E1 凡

Note: U+51E1 凡 would not be annotated to refer to U+2007D 凡.

Group 4

4.1 Annotate 沝 (U+2DC09) with

- archaic form of 6CF2 泝

→ 6CCD 沝

Note: U+6CCD 沝 would not be annotated to refer to U+2DC09 沝.

Conclusion

I would like to request IRG members to give feedback on this idea.

If IRG members support this idea in principle, I would be grateful if more such groups could be suggested.