

Korea JTC1/SC2, Committee on Character Codes

Author: YANG Wangsung; KIM, Kyongsok

Date: 2012.11.08.

Status: National Body Position, Rep. of KOREA

Subject: comments WG2 N4234 and M59.03 RE: PUA chars in IDS

1. Background

R.O.Korea reviewed WG2 N4234 and M59.03 which suggests to use PUA characters in IDS and would like to make some comments

1.1 quoted from Annex I of ISO/IEC 10646

An IDS consists of an IDC followed by a fixed number of Description Components (DC). A DC may be any one of the following:

- a coded ideograph*
- a coded radical*
- another IDS*

NOTE 1 – The above description implies that any IDS may be nested within another IDS.

...

NOTE 2 – An IDS is restricted to no more than 16 characters in length. Also no more than six ideographs and/or radicals may occur between any two instances of an IDC character within an IDS.

1.2 quoted from WG2 N4234 by Unicode and US

ISO/IEC 10646 defines Ideographic Description Sequences (IDSes) in Annex I, and imposes two kinds of restrictions:

- the components can only be coded radicals or coded ideographs
- Note 2 limits the overall length of an IDS, and the distance between successive Ideographic Description Characters in an IDS

==>

The proposal is to modify annex I to:

- add private use characters to the list of possible components
- *remove Note 2.*

1.3 WG2 resolution M59.03

RESOLUTION M59.03 (IDS Syntax in Annex I):

Korea (Republic of) - Abstained

Canada, China, Finland, Ireland, Japan, Lithuania and USA - Accepted

WG2 accepts the proposal from the US national body in document N4234 to remove the deficiency and align the definition of IDS with the current practice, and instructs its project editor to modify the text in Annex I accordingly.

3. R.O.Korea's analysis and suggestion

1) If we use PUA characters in IDS, these PUA characters will be used quite widely among Hanzi users and researchers. Using PUA characters this way seems to have deviated fairly far from the original idea of PUA characters.

2) Furthermore, since most users will not have font(s) including these PUA characters, CJK Unified (main, ExtA/B/C/D) and CJK Compatibility Hanzi characters (and possibly other characters too), they will have difficulty displaying and understanding IDS correctly (i.e., PUA characters will NOT be displayed or other strange PUA characters will be displayed instead).

It was quite a serious problem for R.O.Korea. For example, in preparing Hanja characters for Extension F submission, R.O.Korea tried to check whether these characters are already encoded in UCS. However, we could not see those PUA characters in IDS of existing Hanzi chars and we could not type easily these PUA characters in an IDS of new Hanzi chars.

3) According to WG2 N4241 by Ken Lunde, PUA characters are included in IDS of about 2,600 CJK characters. This number is not small.

4) **conclusion:** Based on the above analysis, R.O.Korea concluded that it is not desirable to use PUA characters in IDS.

5) **suggestion:** It was mentioned that, about 10 years ago, China proposed about 500 component characters which are used only as a component of Hanzi characters but not as an independent Hanzi character.

R.O.Korea suggests that IRG review carefully that proposal once again for possibly adding them to UCS and use for IDS.

* * *