

2021/03/13

Feedback to IRG N2464

suzuki toshiya

Abstract

Basically, I support the points raised by Professor Kim, there are some points to be documented in IRG PnP and Working Document Series. I try to answer a few questions and write a few comments (mainly for the part 1, and partially for the part 2 of IRG N2464).

Comments on Kawabata-san's program

The questions I can try to answer are not so many, so please let me put the number for this document.

Q1: The equivalence / unification candidate program is ready to handle Supp. Comps.?

A1: Yes, but depends on the quality of the maintenance condition of the past ids.txt. Please think about the situation:

IRG MB A submits a character A-0123, whose IDS is 𠄎示&H7-01;

IRG MB B submits a character B-2345, whose IDS is 𠄎禛 &H7-01;

The program by Kawabata-san would find the similarity between A-0123 and B-4567, with no knowledge about the glyph shape of &H7-01;

But, another situation, like,

IRG MB X submits a character X-6789, whose IDS is 𠄎目&H3-04;

The program by Kawabata-san would be hard to find its similarity with 具, because the program and its currently predefined IDS data have no knowledge about the glyph shape of &H3-04;. To improve this situation, we should update the IDS for 具 - currently it has 2 IDSs (𠄎&CDP-8BA5; 八 and 𠄎目一八).

Yet I'm unsure whether addition of the 3rd IDS (𠄎目&H3-04;) is enough, or we should replace old 2 IDSs by new one.

This question is strongly related with the question raised in 2.6 "IDS for URO, ExtA ~ ExtG should be updated?". I think, if we include some entries like a macro symbol definition (of programming language C), like,

&H3-04; 𠄎一八

might improve the situation.

Q2: IRG will propose to modify IDS-relevant clauses in ISO/IEC 10646 to reflect Supp. Comps. in IRG2225?

A2: Maybe it would be dealt with as out of scope of ISO/IEC 10646. ISO/IEC 10646 or Unicode can discuss the plain text without markup. But the syntax of Supp. Comps. is clearly XML-like. Even if we try to write such in ISO/IEC 10646, we would be asked a question “Are you saying this is plaintext? It looks like a portion of the marked-up language, the syntax of the marked-up language does not fit to ISO/IEC 10646”. The better way is trying to find an appropriate organization to define the standard for the marked-up language for the light-weight ideographic structure description, if IRG has to take the responsibility about the usage of IDS data out of IRG.

Comments on other issues

- I agree with the requirement of the update of IRG PnP (or IRG Working Document) to clarify the syntax of IDS which can be used in the submission to IRG.
- I agree with the requirement of the form for IRG MBs to propose/submit new Supp. Comps, to minimize the cost of the consolidation and review works.
- I agree with the requirement of the definition of the rule to name new Supp. Comps used in the working set submissions, to minimize the cost of the review works.
- About a few “new” syntaxes proposed by Kawabata-san (IRG N2464 part 2, 2.11 and 2.12), IRG is expected to issue an official recommendation for the future submissions, which format is the best for IRG. It would prevent the case that the different syntaxes are used in multiple submissions, and it would be useful to reduce the cost of the consolidation & review.
- The feedback from Bai Yi demonstrates the automation of the update of the IDS data after the standardization of new extensions in CJK Unified Ideograph. I wish if it can be used to estimate “how many IDSs could be optimized by the introduction of this new Supp Comp”.

(end of document)