

Universal Multiple-Octet Coded Character Set UCS

ISO/IEC JTC1/SC2/WG2/IRG N1988

Date: 2014-03-21

Source:	China
Title:	Additional Request for the 3 China's UNCs
Meeting:	IRG#42, Qingdao
Actions required	To be considered by IRG
Distribution:	IRG
Medium :	Electronic
Page:	2
References:	IRGN1967 (WG2N4508)

The WG2#62 confirmed China's needs of the 3 urgently needed characters which were proposed in WG2N4508 (IRGN1967), and asked IRG to process them in CJK_F. (See WG2 Resolution M62.15.) Having communicated with WG2 project editor and other WG2 experts, China requests IRG to process the 3 UNCs as below:

1. Add the 3 UNCs to CJK_F1.
2. If CJK_F1 could not be finalized after IRG#42 as scheduled, request WG2 to put the 3 UNCs to the end of CJK block (i.e. U+9FCD, U+9FCE, U+9FCF) in the nearest AMD of UCS Edt.4.

The 3 UNCs:

垚

1.
 - *General Purpose Normalized Hanzi List 6774.*
 - Never proposed before.
 - G source: G_GFZ108

砒

2.

- *General Purpose Normalized Hanzi List* 7146.
- CJK_F Serial Number 05113.
- G source: G_GFC022.

鍤

3.

- *General Purpose Normalized Hanzi List* 7373.
- CJK_F Serial Number 07656.
- G source: G_GFC059.
- **UTC source: UTC-01171**

End.

Subject: Re: IRG N1667 (IRG #34)
From: Ken Lunde <lunde@adobe.com>
Date: 2/4/2014 8:59 PM
To: Lu qin <csluqin@comp.polyu.edu.hk>
CC: chen-zhuang <chenzh-zhuang@163.com>, Michel Suignard
<michel@suignard.com>, Andrew West <andrewcwest@gmail.com>,
Jaemin Chung <jaemin_chung@hotmail.com>, Cendy Li
<cendyli@ogcio.gov.hk>, "sckwong@ogcio.gov.hk"
<sckwong@ogcio.gov.hk>, "skhlaw@ogcio.gov.hk"
<skhlaw@ogcio.gov.hk>

Dr. Lu and others,

As a font developer, and someone who has frequent contact with font developers throughout East Asia, I can chime in (again) with some thoughts, some of which are repeated from previous emails in this discussion.

Anyway, from a "frequency of actual use" point of view, it is correct that the traditional form (as used in Hong Kong SCS) sees much more. Mr. Chen's search results were very predictable.

However, the number of GB 18030-compliant font implementations far outnumber the number of Hong Kong SCS implementation, probably by at least an order or magnitude. GB 18030 compliance represents a barrier to the China market, and this compliance entails support for all of Extension A. It doesn't matter about frequency of usage, but rather that the character is in Extension A.

In other words, it would be much more painful for the font industry if the simplified form would move. There is pain involved with moving the traditional form, but it is far less pain when you consider who would be affected by the process.

My vote would thus be to move the traditional form of U+4CA4, which corresponds to Hong Kong SCS 0x9D73, to U+9FD0.

Regards...

-- Ken

On Apr 2, 2014, at 3:03 AM, Lu qin <csluqin@comp.polyu.edu.hk> wrote:

> > Dear Chen Zhuang,
> >

> > Thank you very much for your response. This is very helpful. Do let us know when you have more information. This might be an over-unification in the super CJK years.

> >

> > Best regards,

> >

> > Lu Qin

> >

> > On 2/4/2014 3:23 PM, chen-zhuang wrote:

>> >> Dear Dr. Lu,

>> >> I searched for U+4CA4 (𠃹) or "𠃹, simplified" on the internet. It seems like that the character is not used in mainland of China. But, the character is specified in GB 18030 which is an important standard, so I am not sure if it is used by some people. I'm asking my colleagues to give me information if they have.

>> >> We are really worrying about moving the simplified one to 9FCx. The character was specified in GB 18030-2000 and GB 18030-2005 (0x8234EB33). Some Chinese fonts vendors such as Founder Group, Jade Bird Huaguang have developed variants types of fonts and other softwares. Besides, lots of Japanese and US companies including Hitachi and Microsoft are following the standard when they are developing or selling products in China. I don't know if the change will impact these products.

>> >> Let me have more information in mainland of China.

>> >> Regards,

>> >> Chen Zhuang

>> >>

>> >>

>> >>

>> >>

>> >>

>> >>

>> >>

>> >>

>> >>

>> >>

>> >>

>> >>

>> >>

>> >> '/2014fí04'-02 14 ±00Σ÷f""Lu
qin"<csluqin@comp.polyu.edu.hk>-¥μ¿fj

>> >>

>> >> Dear Chen Zhuang,

>> >>

>> >> Can you give some feedback on this issue related to the dis-unification

>> >> issue. Michel needs your input before suggesting a solution.

>> >>
>> >> Thanks,
>> >>
>> >> Lu Qin
>> >>
>> >> On 1/4/2014 2:12 PM, Michel Suignard wrote:
>>> >>> Dear Chen Zhuang
>>> >>> I see you have taken actions on the other 3 UNC
characters on another email thread.
>>> >>>
>>> >>> However I'd like to have your opinion on the proposal to
'dis-unify' 4CA4 which was discussed in the same thread.
Especially important is the following consideration:
>>> >>> <<
>>>> >>>> When one considers the number of existing fonts that
include the simplified form at U+4CA4, they are all GB 18030
ones. The original source for U+4CA4 is Singapore, and the IRG
hasn't heard from them for years. So, from a practical point of
view, the simplified form of U+4CA4, though present in a larger
number of fonts, is not used very much. The Hong Kong form,
which uses the traditional form of the left-side radical, is
arguably in more common use, and can be referenced in an actual
national standard (Hong Kong SCS). It is clear that something
should be done, and the sooner it is done, the less painful it
will be, especially for the losing party.
>>>> >>>>
>>> >>> Clearly Hong Kong would prefer 4CA4 to go their way and
a new character to represent the simplified version, likely in
the 9FCx range along with the 3 UNCs (even though the simplified
version was encoded first in 4CA4). The other way (which Hong
Kong likes much less) is to keep the G source at 4CA4 as it is
and encode/move the H source in a new location (in 9FCx).
>>> >>>
>>> >>> I'd like to have your opinion in the matter. Eventually
the WG2 experts and SC2 National Bodies will decide through
ballot but I'd like to have a starting position as close to the
preferred options by the interested parties.
>>> >>>
>>> >>> Best regards,
>>> >>>
>>> >>> Michel
>>> >>>
>>> >>> -----Original Message-----
>>> >>> From: Lu qin [mailto:csluqin@comp.polyu.edu.hk]
>>> >>> Sent: Thursday, March 20, 2014 11:59 PM
>>> >>> To: Ken Lunde; Michel Suignard
>>> >>> Cc: Andrew West; chen-zhuang; Jaemin Chung; Cendy Li;

sckwong@ogcio.gov.hk; skhlaw@ogcio.gov.hk

>>> >>> Subject: Re: IRG N1667 (IRG #34)

>>> >>>

>>> >>> Dear Everyone,

>>> >>>

>>> >>> Sorry that I was absent from this discussion because I need to finish up

>>> >>> something this weekend. My opinion is that HKSARG had made a proposal

>>> >>> in IRGN1667 and it will stand for its position. I will advise HKSARG to make a new request with reference to the previous request for an urgent processing request. This document can then reach to all IRG members and also to Michel is the project editor. Michel can of course "tentatively" propose a solution to bring to the attention of WG2 members. This will then come back as a feedback "action" from Michel as input to IRG No. 42.

>>> >>>

>>> >>> I will assign the document number IRG N1989 to HKSARG for their immediate action. Once I receive that document, we can have the sequence of actions.

>>> >>>

>>> >>> Best regards,

>>> >>>

>>> >>> Lu Qin

>>> >>>

>>> >>>

>>> >>> On 20/3/2014 4:42 AM, Ken Lunde wrote:

>>>> >>>> Michel,

>>>> >>>>

>>>> >>>> Given the additional details that you provided, I completely agree

>>>> >>>> with your disagreeing with me about the timing. ;-)

>>>> >>>>

>>>> >>>> Your suggestion to include the additional character now goes along

>>>> >>>> with "the sooner the better and the less painful" approach. (In fact,

>>>> >>>> given the latitude that you wield, you might even consider going a

>>>> >>>> step further and stick in China's three UNCs, which would effectively

>>>> >>>> short-circuit the possibility of Extension F not being ready to submit

>>>> >>>> to WG2 after IRG #42.)

>>>> >>>>

>>>> >>>> Hopefully Dr. Lu and Mr Chen can chime in with some

form of agreement. In the end, it is an issue that primarily affects Hong Kong SAR in terms of actual usage, and secondarily China and Singapore in terms of history.

>>>> >>>>

>>>> >>>> Regards...

>>>> >>>>

>>>> >>>> -- Ken

>>>> >>>>

>>>> >>>> On Mar 19, 2014, at 1:32 PM, Michel Suignard
<michel@suignard.com> wrote:

>>>> >>>>

>>>>> >>>>> Ken,

>>>>> >>>>> I disagree somewhat on the timing. Amd2 is going in pdam ballot in a month or two (I just need to make sure it is at least 3 months before next WG2 meeting late September. If we did endorse the dis-unification only in September for U+4CA4, that gives just one round of technical ballot (Amd2 is likely to go DAM2 ballot after September).

>>>>> >>>>>

>>>>> >>>>> I'd rather put the new character now in amd2 with one option and let IRG #42 discuss about and give suggestions to its member bodies when they vote on amd2. That gives more time to come to a conclusion. When WG2 meets in September the encoding can always be reversed or even pulled out of amd2 if that's the group decision.

>>>>> >>>>>

>>>>> >>>>> Given the perceived urgency I'd rather make it visible sooner than later. At working group level, the project editor has much more latitude to put things in ballot that has been past habits in WG2. It is only the enquiry phase (DAM or DIS) that requires some formality and an explicit endorsement from SC2 members. With WG2 meeting less and less frequently it is important to use all the flexibility we have.

>>>>> >>>>>

>>>>> >>>>> However I would like to put the option that is likely to win in pdam2. That's why having opinion from our Chinese colleagues would be useful.

>>>>> >>>>>

>>>>> >>>>> Michel

>>>>> >>>>>

>>>>> >>>>> -----Original Message-----

>>>>> >>>>> From: Ken Lunde [mailto:lunde@adobe.com]

>>>>> >>>>> Sent: Wednesday, March 19, 2014 12:15 PM

>>>>> >>>>> To: Michel Suignard

>>>>> >>>>> Cc: Andrew West; chen-zhuang; Lu qin; Jaemin Chung

>>>>> >>>>> Subject: Re: IRG N1667 (IRG #34)

>>>>> >>>>>

>>>> >>>> Michel,

>>>> >>>>

>>>> >>>> When one considers the number of existing fonts that include the simplified form at U+4CA4, they are all GB 18030 ones. The original source for U+4CA4 is Singapore, and the IRG hasn't heard from them for years. So, from a practical point of view, the simplified form of U+4CA4, though present in a larger number of fonts, is not used very much. The Hong Kong form, which uses the traditional form of the left-side radical, is arguably in more common use, and can be referenced in an actual national standard (Hong Kong SCS). It is clear that something should be done, and the sooner it is done, the less painful it will be, especially for the losing party.

>>>> >>>>

>>>> >>>> I agree that it is difficult to decide which should be moved and encoded elsewhere, but in the end, I think that the decision will be based on which character was there first.

>>>> >>>>

>>>> >>>> Extension A became part of Unicode in Version 3.0 (09/1999), and the traditional form became part of Hong Kong SCS in its 2001 version, so the simplified form definitely got there first.

>>>> >>>>

>>>> >>>> Given that the decision cannot happen until WG2 #63, and because IRG #42 takes place before then, I suggest that the IRG take this up and provide a recommendation to WG2.

>>>> >>>>

>>>> >>>> Regards...

>>>> >>>>

>>>> >>>> -- Ken

>>>> >>>>

>>>> >>>> On Mar 19, 2014, at 11:40 AM, Michel Suignard <michel@suignard.com> wrote:

>>>> >>>>

>>>>> >>>>> My only concern is that some lose in that game. All commercial fonts targeting Hong Kong will show the traditional shape for U+4CA4 for the foreseeable future. HKSAR had proposed in IRG N1667 to keep the traditional version at 4CA4 and encode the simplified version somewhere else which is the opposite of what is proposed here. At the same time, nothing prevents the HK fonts to show the same character at 4CA4 and 9FCD for a transitional period if we go the opposite way.

>>>>> >>>>>

>>>>> >>>>> It depends which entity is the biggest user of the current position at 4CA4. Is it GB or Big5? It'd be nice to have Lu Qin opinion. Obviously the original glyph in 4CA4 has always used the simplified radical so it is really in essence a faulty

unification decision from HKSAR when they did HKSCS.

>>>>> >>>>>

>>>>> >>>>> This was brought 4 years ago, why didn't anything happen until now? I can't find any mention of a decision in either minutes and resolutions of that meeting.

>>>>> >>>>>

>>>>> >>>>> Logistically it is breeze to bring this to an amendment (Amd 2 for now). I have a lot of latitude in bringing a character in a proposal (has been reaffirmed now several times in WG2). And all the information about sources and the fonts with the glyphs are available. I don't have to wait for September.

>>>>> >>>>>

>>>>> >>>>> So it is a really a matter to decide which way to go. I understand that going the HKSAR way would require a change in the RSUnicode value for 4CA4. I understand that so far that no one here is proposing to go for what IRG N1667 was proposing. I just want confirmation before proceeding.

>>>>> >>>>>

>>>>> >>>>> Michel

>>>>> >>>>>

>>>>> >>>>> -----Original Message-----

>>>>> >>>>> From: Ken Lunde [mailto:lunde@adobe.com]

>>>>> >>>>> Sent: Wednesday, March 19, 2014 5:53 AM

>>>>> >>>>> To: Andrew West

>>>>> >>>>> Cc: chen-zhuang; Lu qin; Jaemin Chung; Michel Suignard

>>>>> >>>>> Subject: Re: IRG N1667 (IRG #34)

>>>>> >>>>>

>>>>> >>>>> +1

>>>>> >>>>>

>>>>> >>>>> On Mar 19, 2014, at 1:53 AM, Andrew West <andrewcwest@gmail.com> wrote:

>>>>> >>>>>

>>>>>> >>>>>> On 19 March 2014 00:21, Ken Lunde <lunde@adobe.com> wrote:

>>>>>>> >>>>>>> The other way to look at the simplified form of U+4CA4 is that it is part of the required portion of GB 18030, which also suggests encoding the traditional form (with an H source) elsewhere, either appended to the URO as a single-character UNC submission or thrown in with Extension F (F1).

>>>>>>> >>>>>>> I personally think that the sooner the mistaken unification is

>>>>>>> >>>>>>> undone the better, so I would advocate adding the traditional form

>>>>>>> >>>>>>> of U+4CA4 to the space at the end of the basic CJK Unified

>>>>>> >>>>>> Ideographs block (i.e. at U+9FCD) at the
earliest opportunity.
>>>>>> >>>>>>
>>>>>> >>>>>> If the IRG meeting in May can agree to submit a
document to WG2 in
>>>>>> >>>>>> time for September's meeting in Sri Lanka then I
would think that
>>>>>> >>>>>> it could be put straight on the Amd. 2 ballot.
Michel, what do you think?
>>>>>> >>>>>>
>>>>>> >>>>>> Andrew
> >

Subject: RE: Re:RE: Resolution M62.15
From: Michel Suignard <michel@suignard.com>
Date: 4/3/2014 3:49 AM
To: chen-zhuang <chenzh-zhuang@163.com>
CC: kenlunde <lunde@adobe.com>, "csluqin@comp.polyu.edu.hk" <csluqin@comp.polyu.edu.hk>, "mikeksar@10646.com" <mikeksar@10646.com>, "umavs@ca.ibm.com" <umavs@ca.ibm.com>, "dwanders@sonic.net" <dwanders@sonic.net>

Dear Chen

I checked the attribute table and that blows the myth about having F1 and F2 cleanly split at a given value. I see in that spreadsheet that F1 has many high radical values and that F2 has many low radical values. Considering that, I would suggest that we keep the current plan of record (which is option 3 below). If you feel that there is any danger that ExtF could be late (i.e. not being proposed in September in Colombo), we could have an insurance policy by adding these 3 in amd2 at 9FCD-9FCF (option-1) and in Colombo we have the choice to keep there as they are (if ext F is late) or in ext F as currently planned if ext F is on time. Having them balloted before Colombo expands our available options.

It also means though that IRG has to move #5513 and #7656 from F2 to F1 so that they can be formally included in what will become CJK Extension F.

If you want a proposal for them in amendment 2 (even temporarily) I need source reference info (with some descriptions if their dictionary/reference was not used before) and a font with these 3 glyphs. Not urgent because I won't issue pdam2 before late April or even early May but it doesn't hurt to get the info early. I see that the two existing F entries (#5513 and #7656) use G_Z and G_GFC notations which have no obvious connection with existing references.

Michel

From: chen-zhuang [mailto:chenzh-zhuang@163.com]
Sent: Monday, March 3, 2014 3:28 AM

To: kenlunde; csluqin@comp.polyu.edu.hk; mikeksar@10646.com;
Michel Suignard; umavs@ca.ibm.com; dwanders@sonic.net
Subject: Re:RE: Resolution M62.15

Dear all,

Thanks for being concerned about the 3 China's UNC.

Basically, I prefer to the following options:

add the all 3 characters to 9FCD, 9FCE, 9FCF.

add CJK_F 05113 𐤀𐤁 and CJK_F 076560f-Ū to 9FCD, 9FCE, and
Ōiâ to CJK_F which will be finished by IRG in this May.

add all to CJK_F.

For option 3, it is OK too. Actually, there already exist
radical #112 𐤀 and radical #167 in CJK_F1 for Japan's
government-use characters, such as JMJ-058197 𐤀𐤁 and
JMJ-060286 𐤀𐤁. There are more radicals out of the range of
first half of Kangxi radicals. All these characters are kept in
CJK_F1 according to Japan's request, agreed at IRG39. See
attributes table of IRGn1979 CJK_F1v2.0.

Best regards,

Chen Zhung

06 ±23Σ÷f""ken lunde"<lunde@adobe.com>—¥μ¿fj