

修訂《香港增補字符集》字符 H-9B4C 和大五碼字符 0xA2CD 的 ISO/IEC 10646 碼位建議

背景

2015 年 8 月 16 日，中文界面諮詢委員會秘書處收到表意文字小組文件 N2077 “Re-mapping of H-9B4C and HB0-A2CD (kIRG_Hsource)”（[附件一](#)）。該文件提出兩項建議，分別關於修訂《香港增補字符集》字符 H-9B4C 和大五碼字符 0xA2CD 的 ISO/IEC 10646 碼位事宜。秘書處已加以分析，並整理有關資料，供工作小組成員考慮。

修訂字符 H-9B4C 的 ISO/IEC 10646 碼位

2. 文件 N2077 的建議一，是把《香港增補字符集-1999》字符“**槩**” (H-9B4C) 的碼位由 U+69E9（收錄於 ISO/IEC 10646-1:1993 及其後版本）更改為 U+3BA3（收錄於 ISO/IEC 10646-1:2000 及其後版本）。

槩
69E9
69E9

(HKSCS-1999)

3. 下表列出 U+69E9 和 U+3BA3 兩字在 ISO/IEC 10646-1:2000 的字形。U+69E9 包括三款等同字形，其中 E-2E65 的字形與 U+3BA3 相同。這情況亦見於其後的版本（[附件二](#)）。

ISO/IEC 10646-1:2000	105/233 69E9	槩 E-2E65 E-1469	槩 3-4C64 3-4468	槩 1-576B 1-5575
	059/163 3BA3	槩 3-4167 3-3371		

4. ISO/IEC 10646 還有下列多個包含“**槩**”這個部件的字符，當中既有等同字形分別編碼，也有等同字形單一

編碼，但沒有同一字形分別見於兩個碼位的情況。這顯示 U+3BA3 的增收應再作檢討。

<p>等同字形 分別編碼</p>	<p>66A8  暨 暨 暨 <small>日 72.10</small> <small>G0-745F HB1-BA5B T1-6924</small></p> <p>66C1 暨 暨 暨 暨 <small>日 72.12</small> <small>GE-2C7B T3-5123 J0-5A7A K1-5B70</small></p> <hr/> <p>6982  概 概 概 概 <small>木 75.9</small> <small>G0-3845 HB1-B7A7 T1-644D J0-3335</small></p> <p>69EA 概 概 概 概 <small>木 75.11</small> <small>GE-2E66 T3-4C60 J3-7624 K0-4B48</small></p> <hr/> <p>6E89  漑 漑 漑 <small>水 85.9</small> <small>G0-3848 HB1-B540 T1-6046</small></p> <p>6F11 漑 漑 漑 漑 <small>水 85.11</small> <small>GE-3071 T3-4764 J0-5E74 K0-4B49</small></p>
<p>等同字形 單一編碼</p>	<p>5ED0 廐 廐 廐 廐 廐 <small>广 53.11</small> <small>GE-2934 H-9DFB T3-4660 J0-567D K0-4F2A</small></p> <p>6168 慨 慨 慨 慨 慨 慨 <small>心 61.9</small> <small>G0-3F2E HB1-B46E T1-5F35 J0-3334 K0-4B46 V1-5577</small></p> <p>6461 概 概 概 概 概 概 <small>手 64.11</small> <small>G3-4A76 HB2-D962 T2-3C27 J1-4065 K2-3665 V2-8C5A</small></p> <p>8507 蔞 蔞 蔞 蔞 蔞 <small>艸 140.11</small> <small>G3-6A28 HB2-DFBD T2-4662 J1-592F K2-5940</small></p>
<p>單一字形</p>	<p>7A4A 概 概 概 概 <small>禾 115.9</small> <small>G3-5C36 HB2-E345 T2-4C4C V1-6143</small></p> <p>20330 偲 偲 偲 <small>人 9.11</small> <small>UCS2003 GKX-0114.22 T5-307B</small></p>

既
69E9
69E9

5. 字符合 ISO/IEC 10646 的等同原則，早在 1999 年已獲納入《香港增補字符集-1999》，並得到業界的支援（支援此字形的字體見下表）。因此，U+69E9 應保留這個字形，以維護用戶資料的兼容性和完整性。

Microsoft YaHei 既	細明體_ HKSCS 既	SimSun 既
Arial Unicode MS 既	SimHei 既	CCCLKaiUT 既

修訂字符 0xA2CD 的 ISO/IEC 10646 碼位

6. 文件 N2077 的建議二涉及三個大五碼符號“十” (0xA2CC)、“卅” (0xA2CD)和“卅” (0xA2CE)配對 ISO/IEC 10646 碼位的事宜。

7. 大五碼字符配對 ISO/IEC 10646 碼位的資料見以下 Unicode 文件：

<http://www.unicode.org/Public/UCD/latest/ucd/Unihan.zip>（內含的 Unihan_IRGSources.txt）

現時 Unihan_IRGSources.txt 只列明 0xA2CD 配對 U+5344 碼位，並沒有包括 0xA2CC 和 0xA2CE 的配對。

8. 建議二具體提議加入“十” (0xA2CC)和“卅” (0xA2CE)與“十” (U+3038)和“卅” (U+303A)的配對；以及把 0xA2CD 的碼位由“卅” (U+5344)更改為“卅” (U+3039)。

9. 大五碼字符“十”(0xA2CC)、“廿”(0xA2CD)和“卅”(0xA2CE)為符號區內的符號，另外在常用字區也有三個中文用字“十”(0xA451)、“卅”(0xA47B)和“卅”(0xA4CA)。這六個字符都不是《香港增補字符集》字符。

大五碼符號區 (A140-A2CE)	A2C0 Ⅷ Ⅸ X 卍 卍 又 8 十 廿 卅 A
大五碼常用字區 (A440-C67E)	A450 比 十 卜 又 三 下 丈 上 丫 九 凡 久 么 也 乞 于 A460 亡 无 刃 勻 千 又 口 土 土 夕 大 女 子 子 丕 寸 A470 小 尢 尸 山 川 工 己 己 巳 巾 干 卅 弋 弓 才 A4A0 丑 丐 丕 中 丰 丹 之 尹 予 云 井 互 五 亢 仁 A4B0 什 行 仆 仇 仍 今 介 仄 元 允 內 六 兮 公 亢 凶 A4C0 分 切 刈 勻 勾 勿 化 匹 午 升 卅 卞 厄 友 及 反

10. 根據“Additional Suzhou numerals”（見 ISO/IEC 10646:2014 第 370 頁），大五碼 0xA2CC 至 0xA2CE 分別配對 U+3038 至 U+303A 是合理的做法。

Additional Suzhou numerals

3038 十 HANGZHOU NUMERAL TEN

≈ 5341 十

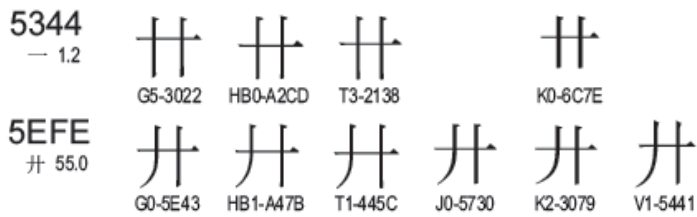
3039 卅 HANGZHOU NUMERAL TWENTY

≈ 5344 卅

303A 卅 HANGZHOU NUMERAL THIRTY

≈ 5345 卅

不過，現時 Unihan_IRGSources.txt 把三個大五碼常用字“十”(0xA451)、“卅”(0xA47B)和“卅”(0xA4CA5)分別配對“十”(U+5341)、“卅”(U+5EFE)和“卅”(U+5345)， “Additional Suzhou numerals”則把符號區的“卅”關聯到文字區的“卅”(U+5344)而非“卅”(U+5EFE)。因此，表意文字小組應考慮一併修訂“Additional Suzhou numerals” U+3039 的關聯，即把 U+3039≈U+5344 改為 U+3039≈U+5EFE。



建議

11. 關於文件 N2077 的建議一，秘書處的意見是提議表意文字小組應慎重考慮用戶資料的兼容性和完整性，讓 U+69E9 保留“𠂇”字形。工作小組可待表意文字小組討論和處理 U+69E9 和 U+3BA3 兩字包含同一字形的情況後，再跟進討論結果。

12. 關於建議二 (0xA2CC 至 0xA2CE 分別配對 U+3038 至 U+303A)，秘書處認為合理，應予支持。

13. 秘書處會收集工作小組成員對上述建議的意見，整理後提交表意文字小組參考。

徵詢意見

14. 歡迎工作小組成員就文件內容發表意見。

政府資訊科技總監辦公室
2015年 8月

Universal Multiple-Octet Coded Character Set
UCS

ISO/IEC JTC1/SC2/WG2/IRG N2077

Date: 2015-08-14

Proposed By	Henry Chan, Ken Lunde, Michel Suignard
Title	Re-mapping of H-9B4C and HB0-A2CD (kIRG_HSource)
Meeting	IRG #44
Status	Individual Contribution
Action	For Consideration by the IRG (Discussion primer: IRGN 2074)
Pages	3

Hong Kong is planning to issue the Hong Kong Character Set (HKCS) in 2015 to replace the existing Hong Kong Supplementary Character Set (HKSCS) standard and migrate away from dependencies on Big-5. However, several mapping issues have been identified.

Issue #1

The form for H-9B4C (see Figure 1.1) specified in the HKSCS standard uses the modern variant of 既 as its top component, which is different from most of the other forms mapped to U+69E9 (see Figure 1.2). The form used would more appropriately be mapped to U+3BA3 (see Figure 1.3).

既
木


Figure 1.1 – Form for H-9B4C, currently mapped to U+69E9

69E9 木 75.10	既	既	既	既	既	既
	GE-2E65	H-9B4C	T3-4C64	J3-7623	K1-576B	V2-8D3A


Figure 1.2 – ISO/IEC 10646 Code Chart for U+69E9

3BA3 木 75.9	既	既
	GHZ-21257.01	T3-4167

Figure 1.3 – ISO/IEC 10646 Code Chart for U+3BA3

Compatibility is not an issue. Currently the default fonts for the Windows and OS X operating systems in use in Hong Kong display the T-Source form for U+69E9. Choosing the character  in an input method editor would result in U+3BA3. Furthermore, as the Hong Kong SAR government strongly encourages the use of ISO/IEC 10646 in computer systems, round-trip mapping would not be a problem.

When the Hong Kong Character Set 2015 (HKCS) is released, which will comprehensively cover characters in use in Hong Kong, computer vendors may finally release fonts that follow the relevant glyph standards. However, if that happens, the glyph for characters encoded using U+69E9 would change unexpectedly. Since U+69E9 and U+3BA3 would display as the same glyph, users would also be unable to produce the traditional variant, which is not preferred by the forthcoming HKCS standard.

The Chinese Language Interface Advisory Committee (CLIAC) of the Office of the Chief Information Officer, HKSARG has discussed changing the glyph at U+69E9/H-9B4C to use the traditional variant of  as its top component, similar to the T-Source of U+69E9. However, it was concluded that the glyph shape for H-9B4C could not be modified. Therefore, the only remaining option is to re-map H-9B4C to U+3BA3.

Re-map Precedent

Corrections to HKSCS mappings have been done in the past. Per WG2 N4621, the H-Source for U+3D1D was changed to U+2A3ED, to better match the glyphs for other regions. The proposal was accepted by WG2 per WG2 N4604 Recommendation M63.05.

Recommendation

Re-map H-9B4C to U+3BA3.

Issue #2

HB0-A2CD is currently mapped to U+5344. However, in Big-5, 0xA2CD is grouped with other Suzhou Numerals. It should instead map to U+3039 HANGZHOU NUMERAL TWENTY instead. Also, the currently unmapped characters 0xA2CC and 0xA2CE should also be mapped to U+3038 and U+303A, respectively.

The mapping omission has likely arisen due to U+3038 HANGZHOU NUMERAL TEN, U+3039 HANGZHOU NUMERAL TWENTY, and U+303A HANGZHOU NUMERAL THIRTY being added to the standard at a later point in time (ISO/IEC 10646-1:2000).

The T-Source for U+5344 has been updated from T1-243F to T3-2138. T1-243F corresponds to Big-5 code point 0xA2CD.

Summary

































Big-5 Code Point	Current UCS Mapping	Recommended UCS Mapping
0xA2C3 	U+3021 HANGZHOU NUMERAL ONE	NO CHANGE
0xA2C4 	U+3022 HANGZHOU NUMERAL TWO	NO CHANGE
...
0xA2CA ≡	U+3028 HANGZHOU NUMERAL EIGHT	NO CHANGE
0xA2CB ㄨ	U+3029 HANGZHOU NUMERAL NINE	NO CHANGE
0xA2CC 卅	U+FFFF (Unmapped)	U+3038 HANGZHOU NUMERAL TEN
0xA2CD 卅	U+5344 CJK UNIFIED IDEOGRAPH-5344	U+3039 HANGZHOU NUMERAL TWENTY
0xA2CE 卅	U+FFFF (Unmapped)	U+303A HANGZHOU NUMERAL THIRTY

Suggestion

Re-map HB0-A2CD to U+3039, map HB-A2CC to U+3038, and map HB-A2CE to U+3040. For backward compatibility, some implementers may choose to map both code points (U+3039 and U+5344) to the same glyph.

- End of document

U+69E9 和 U+3BA3 兩字在 ISO/IEC 10646 各版本的字形

版本	字形
ISO/IEC 10646-1:2000	<p>105/233 69E9   </p> <p>E-2E65 3-4C64 1-576B E-1469 3-4468 1-5575</p> <p>059/163 3BA3 </p> <p>3-4167 3-3371</p>
ISO/IEC 10646:2003	<p>105/233 69E9   </p> <p>E-2E65 3-4C64 1-576B E-1469 3-4468 1-5575</p> <p>059/163 3BA3 </p> <p>3-4167 3-3371</p>
ISO/IEC 10646:2011	<p>69E9 木 75.10      </p> <p>GE-2E65 T3-4C64 J3-7623 K1-576B V2-8D3A H-9B4C</p> <p>3BA3 木 75.9  </p> <p>GHZ-21257.01 T3-4167</p>
ISO/IEC 10646:2012	<p>69E9 木 75.10      </p> <p>GE-2E65 H-9B4C T3-4C64 J3-7623 K1-576B V2-8D3A</p> <p>3BA3 木 75.9  </p> <p>GHZ-21257.01 T3-4167</p>
ISO/IEC 10646:2014	<p>69E9 木 75.10      </p> <p>GE-2E65 H-9B4C T3-4C64 J3-7623 K1-576B V2-8D3A</p> <p>3BA3 木 75.9  </p> <p>GHZ-21257.01 T3-4167</p>