

HKCS 中一些符號的大五碼和 U+碼的映射建議

1 背景

在建立《香港字符集》（HKCS）時，應包括所有大五碼中支持的符號。在審視大五碼的符號時，我們發現有些符號的大五碼和 U+碼¹之間的映射存在問題，其中包括：（1）大五碼和 U+碼存在多對一的映射；（2）大五碼和 U+碼的映射缺失（No Mapping）；（3）大五碼和 U+碼的映射存在疑問；（4）大五碼中原本無歐元符號。因此，需要討論在 HKCS 中如何處理這些問題。

2 問題一：大五碼和 U+碼存在多對一的映射

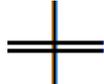
2.1 問題描述

以下八個符號，大五碼和《香港增補字符集》（HKSCS）有重複編碼，而該符號在 ISO/IEC 10646 只存在一個碼位，因此從大五碼到 ISO/IEC 10646 存在 2-1 映射，也就是說，在從 U+碼轉回大五碼時，對所轉的大五碼碼位要有定奪。以下為相關的字符和碼位：

U+碼	大五碼 ²	HKSCS
 U+256D	 HB-A27E	 H-F9FA
 U+256E	 HB-A2A1	 H-F9FB
 U+2570	 HB-A2A2	 H-F9FC
 U+256F	 HB-A2A3	 H-F9FD

¹ 為方便描述，將 ISO/IEC 10646 的編碼簡稱“U+碼”。

² 大五碼的符號圖形來源於：《電腦用中文字型與字碼對照表》（技術通報 C-26），財團法人資訊工業策進會，1984 年。

U+碼	大五碼 ²	HKSCS
 U+2550	 HB-A2A4	 H-F9F9
 U+255E	 HB-A2A5	 H-F9E9
 U+256A	 HB-A2A6	 H-F9EA
 U+2561	 HB-A2A7	 H-F9EB

2.2 建議的處理方法

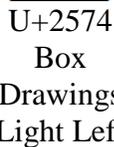
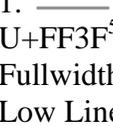
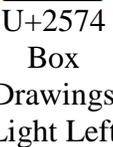
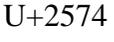
參考 Microsoft 的方法，我們建議 HKCS 使用和 Microsoft 同樣的解決方式，使系統的兼容性最好。Microsoft 的處理方法是：

- (1) 從大五碼轉換為 U+碼時，原本的大五碼“A2XX”和 HKSCS 中的碼位“F9XX”對應到相同的 U+碼“U+25XX”。
- (2) 從 U+碼轉回大五碼時，四個 U+碼對應到原本的大五碼“A2XX”，另外四個 U+碼對應到 HKSCS 中的碼位“F9XX”，詳細信息如下：

U+碼	HKCS
U+2550	H-F9F9 (HKSCS)
U+255E	H-F9E9 (HKSCS)
U+2561	H-F9EB (HKSCS)
U+256A	H-F9EA (HKSCS)
U+256D	HB-A27E
U+256E	HB-A2A1
U+256F	HB-A2A3
U+2570	HB-A2A2

3 問題二：大五碼和 U+碼的映射缺失 (No Mapping)

以下五個大五碼符號，在 Unicode 初始的 Mapping Table³中對應的 U+碼為“U+FFFD”，表示該組符號當時沒有對 U+碼的映射。參考 Microsoft、Apple 和台灣 CNS11643 的處理方法，並考慮了 Dr. Lunde (Adobe) 的意見，我們建議明確該組符號的映射如下：

大五碼	Microsoft Mapping	Apple Mapping ⁴	台灣 CNS11643 Mapping	Dr. Lunde 的建議	我們的建議
 HB-A15A	 U+2574 Box Drawings Light Left	1.  U+FF3F ⁵ Fullwidth Low Line 2. U+F87F ⁶	 U+2574 Box Drawings Light Left	大五碼 A15A 似乎和 A1C5 () 重複，A1C5 又和 A1C4 () 重複	 U+2574 Microsoft 中： • 大五碼 A1C5 () 對應 U+02CD ( Modifier Letter Low Macron)； • 大五碼 A1C4 () 對應 U+FF3F ( Fullwidth Low Line)

³ 見：<http://www.unicode.org/Public/MAPPINGS/OBSOLETE/EASTASIA/OTHER/BIG5.TXT>

⁴ 本文件中採用 Apple OS X (版本號 10.9.5) 中實際使用的映射關係，有些字符的映射關係和 Unicode 發佈的 Apple Mapping Table (<ftp://ftp.unicode.org/Public/MAPPINGS/VENDORS/APPLE/CHINTRAD.TXT>) 不同。

⁵ Apple 的處理方法為：大五碼 A15A、A1C5、A1C4 都對應 U+FF3F，A15A、A1C5 加不同的 transcoding coding hints，A1C4 不加。

⁶ Apple 的 Mapping Table 中，U+F87C-7F 為 transcoding coding hints，與 Unicode 碼組合，詳細說明見：<ftp://ftp.unicode.org/Public/MAPPINGS/VENDORS/APPLE/CHINTRAD.TXT>

大五碼	Microsoft Mapping	Apple Mapping ⁴	台灣 CNS11643 Mapping	Dr. Lunde 的建議	我們的建議
 HB-A1C3	 U+FFE3 Fullwidth Macron	1.  U+203E Overline 2. U+F87C	 U+FFE3 Fullwidth Macron	 U+FFE3 Fullwidth Macron	 U+FFE3
 HB-A1C5	 U+02CD Modifier Letter Low Macron	1.  U+FF3F Fullwidth Low Line 2. U+F87C	 U+02CD Modifier Letter Low Macron	對應U+FF3F () 似乎更好，但可能大五碼 A1C4 () 已經對應U+FF3F	 U+02CD Microsoft 中，大五碼 A1C4 () 對應 U+FF3F ( Fullwidth Low Line)
 HB-A1FE	 U+FF0F Fullwidth Solidus	1.  U+FF0F Fullwidth Solidus 2. U+F87F	 U+FF0F Fullwidth Solidus	大五碼A1FE似乎和A241 () 重複，只能將大五碼A1FE和大五碼A241都對應U+FF0F	 U+FF0F Microsoft 中，大五碼A241 () 對應 U+2215 ( Division Slash)

大五碼	Microsoft Mapping	Apple Mapping ⁴	台灣 CNS11643 Mapping	Dr. Lunde 的建議	我們的建議
HB-A240	U+FF3C Fullwidth Reverse Solidus	1. U+FF3C Fullwidth Reverse Solidus 2. U+F87F	U+FF3C Fullwidth Reverse Solidus	大五碼A240似乎和A242 () 重複，只能將大五碼A240和大五碼A242都對應U+FF3C	U+FF3C Microsoft 中，大五碼 A242 () 對應 U+FE68 (Small Reverse Solidus)

4 問題三：大五碼和 U+碼的映射存在疑問

以下符號的大五碼和 U+碼的映射有待明確，因為 Unicode 的初始編碼⁷和 Microsoft 或台灣 CNS11643 對應的 U+碼不同，參考 Microsoft、Apple 和台灣 CNS11643 的處理方法，並考慮了 Dr. Lunde (Adobe) 的意見，我們建議明確該組符號的映射如下：

大五碼	Unicode 初始編碼	Microsoft Mapping	Apple Mapping	台灣 CNS11643 Mapping	Dr. Lunde 的建議	我們的建議
HB-A145	U+2022 Bullet	U+2027 Hyphenation Point	U+2027 ⁸ Hyphenation Point	U+2027 Hyphenation Point	我認為 U+2022 更合適，但很多情況是對應至 U+2027、同時支持兩個 U+碼。	U+2027

⁷ 見：<http://www.unicode.org/Public/MAPPINGS/OBSOLETE/EASTASIA/OTHER/BIG5.TXT>

⁸ 本文件中採用 Apple OS X (版本號 10.9.5) 中實際使用的映射關係，Unicode 發佈的 Apple Mapping Table 中，對應 U+00B7 (Middle Dot)。

大五碼	Unicode 初始編碼	Microsoft Mapping	Apple Mapping	台灣 CNS11643 Mapping	Dr. Lunde 的建議	我們的 建議
 HB-A14E	 U+FF64 Halfwidth Ideographi c Comma	 U+FE51 Small Ideographi c Comma	1.  U+3001 Ideographic Comma 2. U+F87D	 U+FE51 Small Ideographi c Comma	 U+FE51 Small Ideographi c Comma	 U+FE51
 HB-A156	 U+2013 En Dash	 U+2013 En Dash	 U+2013 En Dash	 U+2015 Horizontal Bar	 U+2013 En Dash	 U+2013
 HB-A1E3	 U+223C Tilde Operator	 U+FF5E Fullwidth Tilde	 U+FF5E ⁹ Fullwidth Tilde	 U+FF5E Fullwidth Tilde	 U+FF5E Fullwidth Tilde	 U+FF5E
 HB-A1F2	 U+2641 Earth	 U+2295 Circled Plus	 U+2295 Circled Plus	 U+2295 Circled Plus	 U+2295 Circled Plus	 U+2295
 HB-A1F3	 U+2609 Sun	 U+2299 Circled Dot Operator	 U+2609 Sun	 U+2299 Circled Dot Operator	 U+2299 Circled Dot Operator	 U+2299
 HB-A241	 U+FF0F Fullwidth Solidus	 U+2215 Division Slash	 U+FF0F Fullwidth Solidus	 U+2215 Division Slash	 U+FF0F Fullwidth Solidus	 U+2215 (半角)

⁹本文件中採用 Apple OS X (版本號 10.9.5) 中實際使用的映射關係，Unicode 發佈的 Apple Mapping Table 中，對應 U+223C (Tilde Operator)。

大五碼	Unicode 初始編碼	Microsoft Mapping	Apple Mapping	台灣 CNS11643 Mapping	Dr. Lunde 的建議	我們的 建議
 HB- A242	 U+FF3C Fullwidth Reverse Solidus	 U+FE68 Small Reverse Solidus	 U+FF3C Fullwidth Reverse Solidus	 U+FE68 Small Reverse Solidus	 U+FF3C Fullwidth Reverse Solidus	 U+FE68
 HB- A244	 U+00A5 Yen Sign	 U+FFE5 Fullwidth Yen Sign	 U+FFE5 ¹⁰ Fullwidth Yen Sign	 U+FFE5 Fullwidth Yen Sign	如果用全角，對應 U+FFE5。兩個都支持也是一種選擇。	 U+FFE5
 HB- A246	 U+00A2 Cent Sign	 U+FFE0 Fullwidth Cent Sign	 U+FFE0 ¹¹ Fullwidth Cent Sign	 U+FFE0 Fullwidth Cent Sign	同上。如果用全角，對應 U+FFE0。	 U+FFE0
 HB- A247	 U+00A3 Pound Sign	 U+FFE1 Fullwidth Pound Sign	 U+FFE1 ¹² Fullwidth Pound Sign	 U+FFE1 Fullwidth Pound Sign	同上。如果用全角，對應 U+FFE1。	 U+FFE1
 HB- A1C2	 U+203E Overline	 U+00AF ¹³ Macron	 U+203E Overline	 U+203E Overline		 U+203E

¹⁰ Unicode 發佈的 Apple Mapping Table 中，對應 U+00A5 (Yen Sign)。

¹¹ Unicode 發佈的 Apple Mapping Table 中，對應 U+00A2 (Cent Sign)。

¹² Unicode 發佈的 Apple Mapping Table 中，對應 U+00A3 (Pound Sign)。

¹³ U+00AF 是半角字符，所以 HKCS 不跟從 Microsoft。

5 問題四：大五碼中原本無歐元符號

大五碼中原本無 A3E1（歐元符號），在 Microsoft 於 2000 年發行的 Windows ME 中，首度增加了該符號。由於 HKSCS 中無歐元符號，我們建議在 HKCS 中增加該符號，其對應的 U+碼可參考 Microsoft、Apple 和台灣 CNS11643 的處理方法。詳細信息如下：

Microsoft Mapping	Apple Mapping	台灣 CNS11643 Mapping	我們的建議	HKCS
€ U+20AC Euro Sign	€ U+20AC Euro Sign	€ U+20AC Euro Sign	€ U+20AC Euro Sign	€ HD-20AC