香港增補字符集—2016

(擬稿)

香港特別行政區政府

政府資訊科技總監辦公室
與
公務員事務局法定語文事務部
合編

2017 年 4 月

目錄

序言

第一部分　概論 ..... 1 - 1

第二部分　《香港增補字符集—2016》編碼方案 ..... 2 - 1

第三部分　ISO/IEC 10646 架構下的《香港增補字符集—2016》.... 3 - 1

附表一:《香港增補字符集—2016》字符表 ..... i - 1

附表二:《香港增補字符集—2016》新增收的字符 ..... ii - 1

附表三:《香港增補字符集—2016》內的兼容字符 ..... iii - 1

序言

《香港增補字符集》在 1999 年初版，其後三度更新。《香港增補字符集—2001》、《香港增補字符集—2004》和《香港增補字符集—2008》三個版本分別增收 116、123 和 68 個字符，令《香港增補字符集》所收錄的字符數目增加至 5 009 個，為推廣採用 ISO/IEC 10646 國際編碼標準的工作奠定基礎，而《香港增補字符集》也獲資訊科技業界和市民廣泛支持和採用。

ISO/IEC 10646 國際編碼標準由國際標準化組織(International Organization for Standardization, ISO)制訂，是應用於儲存和交換電子資料的國際編碼標準，涵蓋各種主要語文的字符，包括繁體和簡體中文字。ISO/IEC 10646:2014 國際編碼標準已包含逾 80 000 個漢字，為電腦平台提供整體而全面的支援。然而，要建立一個支援全部八萬多字的字集，不僅會令開發時間和費用增加，實際應用上也沒有必要。因此，字體或輸入法軟件開發商會因應個別國家和地區的用字要求，選取適量字符開發支援產品。

《香港增補字符集—2016》增收的字符都已收錄於 ISO/IEC 10646 國際編碼標準。ISO/IEC 10646 國際編碼標準除了增收該字集尚未收錄而香港特別行政區(香港特區)有需要使用的字符外(即以縱向擴展形式加入新字符)，還會為字集內已收錄的字符註明哪些是香港用字，並加入香港字形，藉此反映香港實際使用漢字的情況(即橫向擴展，也就是只提供有關字符的資訊，說明是香港用字並加入字形，而不是新增字符)。這不但方便廠商支援香港實際使用的漢字、發展相關的本地化技術，更可令開發成本和時間減少，有利資訊行業發展更多切合香港需要的產品。

與《香港增補字符集—2008》比較，《香港增補字符集—2016》主要增收了 23 個漢字和 1 個符號，藉此進一步滿足本地中文電子通訊的需要。

<u>編碼方案為電子通訊的基礎</u>

把資料輸入電腦和進行電子通訊時，都必須根據某種預先定義的編碼方案予以編碼。就處理中文資料而言，早在九十年代已有不同的編碼方案，包括大五碼、國標碼、ISO/IEC 10646 國際編碼標準。當時的編碼方案未能涵蓋所有香港常用的字符，電腦用戶要在電腦上使用這些字符時，只得自行在使用者造字區為這些字符編配碼位。這些字符包括人名、地名和粵方言用字。這種做法對獨立運作的電腦而言，並無不妥，但電腦一旦互相連接，在通訊和交換資料的時候，用戶自行定義的造字便會引起資料失誤的問題。

共通中文界面

香港特別行政區政府(政府)“數碼 21”資訊科技策略的措施之一，是為香港特區制訂一套開放和共通的中文界面，令中文電子通訊更加準確。共通中文界面的重點就是全面採用 ISO/IEC 10646 國際編碼標準。

《香港增補字符集》的發展

1995 年，當時的香港政府為方便各部門以電子方式通訊，建立了一套名為《政府通用字庫》的中文字符集，在協調香港用戶為中文定義造字的工作上，邁出了第一步。《政府通用字庫》可補充大五碼基本字的不足，因此深受市民歡迎。後來，政府與中文界面諮詢委員會(中諮會)合作，修訂《政府通用字庫》。中諮會的委員包括學術界、語言學界、資訊科技界和出版界代表。經修訂的字符集收納了從社會各界收集得來的字符，並於 1999 年 9 月正式公布，取名《香港增補字符集》(為了與之後發表的擴充版命名方式一致，也稱為《香港增補字符集—1999》)，共收錄 4 702 個字符，供社會人士共用。

鑑於市民和政府部門不時有需要把新字符納入《香港增補字符集》，在政府與中諮會的共同努力下，《香港增補字符集》字符增收程序和原則在

2000年4月公布。中諮會定期舉行會議，審核《香港增補字符集》的字符增收申請。通過審核而增收的字符會提交國際標準化組織轄下的表意文字小組(Ideographic Rapporteur Group)考慮，以期納入ISO/IEC 10646國際編碼標準。

《香港增補字符集》其後三度更新，所收錄的字符數目增加至5 009個。這四個版本的《香港增補字符集》每個都有兩套編碼方案，一套適用於當時在香港使用的大五碼系統，另一套適用於ISO/IEC 10646國際編碼標準平台。這是為逐步由大五碼系統過渡至ISO/IEC 10646系統的安排。

隨着支援ISO/IEC 10646的平台及產品日漸普及，政府在2008年4月修訂《香港增補字符集》的字符增收原則。自此以後，新增收的字符只會獲編配ISO/IEC 10646碼位，而不會再獲編配大五碼碼位。與此同時，ISO/IEC 10646:2011已收納所有《香港增補字符集—2008》的字符，《香港增補字符集—2008》與ISO/IEC 10646國際編碼標準已完全接軌，完成了全面採納ISO/IEC 10646國際編碼標準的重要一步。

為了令ISO/IEC 10646國際編碼標準更能反映香港實際使用漢字的情況和需要，中諮會在2015年12月再修訂了《香港增補字符集》字符增收程序和原則，並開始編訂《香港增補字符集—2016》。

《香港增補字符集—2016》的編碼方案和字符表詳載於本文件內。《香港增字符集—2016》的相關對應表、《香港增補字符集》字符增收程序和字符增收原則等文件可在以下“共通中文界面”網站下載：
https://www.ogcio.gov.hk/tc/business/tech_promotion/ccli/hkscs/ 。

鳴謝

編訂《香港增補字符集—2016》的工作得以順利完成，實有賴中諮會鼎力協助。中諮會現任委員包括文映霞博士、李家駒博士、秦德超先生、張穎之女士、梁慧珠女士、章濤先生、陸勤教授，M.H.、馮淑儀博士、黃家偉先生、鄧佩玲博士和簡錦源先生，前任委員包括方健僑博士、招亮輝先生、梁崇榆先生、熊世賢先生和謝達安先生，我們謹此致謝。

此外，由陸勤教授擔任召集人的“中文界面諮詢委員會工作小組”和香港理工大學陸勤教授團隊的熊丹女士和尤盛先生，在編訂《香港增補字符集—2016》方面不遺餘力，謹此一併致謝。工作小組的組員除了中諮會的部分委員外，還包括李健康先生、張群顯博士、黃耀堃教授、黎達橋先生、蘭蓀博士和教育局的代表。

第一部分 概論

1.1 本文件旨在提供《香港增補字符集—2016》所收錄的字符和對應的 ISO/IEC 10646 國際編碼標準碼位，並說明《香港增補字符集—2016》在 ISO/IEC 10646 國際編碼標準中的整體編碼架構。《香港增補字符集—2016》與以往公布的《政府通用字庫》和各版本的《香港增補字符集》完全兼容。

1.2 《香港增補字符集—2016》收錄 5 033 個字符，包括 5 009 個《香港增補字符集—2008》原有的字符，以及新增收的 23 個漢字和 1 個符號。

1.3 《香港增補字符集—2016》屬編碼字符集，而非字形標準，當中所列漢字的字形應以《香港電腦漢字參考字形》所載者為依歸。《香港電腦漢字參考字形》可於以下網址下載：
https://www.ogcio.gov.hk/tc/business/tech_promotion/ccli/download_area/ 。

1.4 本文件所用的一些名詞及其定義如下：

名詞	定義
基本多文種平面 (平面 0) Basic Multilingual Plane (BMP, Plane 0)	ISO/IEC 10646 國際編碼標準架構內的基本多文種編碼平面，簡稱基本平面，也稱“平面 0”，碼位由 0000 至 FFFF。
字區 Block	一個編碼區域，所對應的字符具有共同特徵。
字符 Character	用於組織、控制或表示數據的元素之一。

名詞	定義
字形 Character Glyph	在 ISO/IEC 10646 國際編碼標準中，特指構成漢字的可識別抽象圖形符號，獨立於任何實際圖形，為實際圖形的抽象表示。構成漢字字形的要素是筆畫、漢字部首和部件，以及其相對位置。
字符集 Character Set	一組已定義的字符。
中日韓表意文字兼容字符區 CJK Compatibility Ideographs	位於基本多文種平面(平面 0)，為兼容中日韓表意文字區而設，收納在原字集已獲編配不同碼位的異體字符或重複的字符，碼位由 F900 至 FAFF。在 ISO/IEC 10646 國際編碼標準內，這些異體字符與其對應的標準字符等同，但在收錄時已在其原字集內獲編配兩個碼位。因此，ISO/IEC 10646 國際編碼標準建立了這個特殊的編碼兼容區，用以確保中日韓表意文字區不會出現一個字符多個碼位的情況，同時讓原字集內的字符與 ISO/IEC 10646 國際編碼標準內的字符可以雙向轉換，確保反向兼容。兼容區內每個字符都有其對應的中日韓表意文字區編碼字符。
中日韓表意文字兼容字符擴展區 CJK Compatibility Ideographs Supplement	該區為中日韓表意文字兼容字符區的擴展區，位於增補表意文字平面(平面 2)，碼位由 2F800 至 2FA1F。本文件把兼容區和兼容擴展區的漢字簡稱為“兼容字符”。

名詞	定義
中日韓表意文字基本區 CJK Ideographs Main Block	ISO/IEC 10646 國際編碼標準中首個為統一表意文字，包括中日韓表意文字而設的字區，碼位由 4E00 至 9FFF。
中日韓原字集 CJK Ideograph Source	ISO/IEC 10646 國際編碼標準內的中日韓表意文字是在整合中國、日本、韓國及其他國家和地區原有的電腦編碼標準內的字符後得出的。這些原有的電腦編碼標準或規範稱為原字集。原字集的頒布國家和地區都有代號：中國內地(G)、香港(H)、日本(J)、韓國(K)、新加坡(S)、台灣(T)和越南(V)。
中日韓統一表意文字 CJK Unified Ideographs	來自中國、日本、韓國及其他國家和地區原有字符標準內的表意文字經等同規則整合後，獲納入 ISO/IEC 10646 國際編碼標準，這些表意文字統稱為“中日韓統一表意文字”。由於該標準最早版本的表意文字主要來自中國、日本和韓國，因此以“中日韓”為名稱，並沿用至今。這些表意文字在本文也簡稱為“漢字”。
碼位 Code Point	以十六進制方式編配予每個字符的編碼。
編碼字符集 Coded Character Set	按照沒有歧義的規則建立的字符集，並訂明其內每一字符的對應碼位。

名詞	定義
中日韓表意文字擴展區 CJK Ideographs Extension Blocks	指設立中日韓表意文字基本區之後逐漸擴充的中日韓表意文字區，擴展區 A 設於基本多文種平面(平面 0)，其他的擴展區設於增補表意文字平面(平面 2)。
政府通用字庫 Government Common Character Set (GCCS)	1995 年由當時的香港政府建立的一套中文字符集，供政府內部交換和處理中文資料之用。
橫向擴展 Horizontal Extension	為 ISO/IEC 10646 國際編碼標準已收納的字符增補某國家或地區所用字形和字源資料的方式。
H-列 H-Column	中日韓統一表意文字區內每個碼位都有多個字形，並以整列形式列出，旨在支持和定義某一個國家或地區所用的字符。香港使用的漢字載於 H-列。
表意文字 Ideograph	指書寫系統主要是表“意”而非表“聲”的文字，漢字屬於表意文字。
ISO/IEC 10646	國際標準化組織制訂的一套編碼標準，目的是編訂包含所有主要語文的單一字符集。
字源資料 Source Reference	ISO/IEC 10646 國際編碼標準內的中日韓表意文字來自多個不同的國家和地區的字源標準，字源資料就是有關字源標準的索引。

名詞	定義
增補表意文字平面 (平面 2) Supplementary Ideographic Plane (SIP, Plane 2)	根據 ISO/IEC 10646 編碼架構，“平面 2”用作增收中日韓表意文字，碼位由 20000 至 2FFFF。
等同 Unification	把一個碼位編配給兩個或以上圖形看來不同但實際上是異體字符關係的漢字符號。各有關符號代表同一數據元素，因此只選擇其中一個字形作為代表。
縱向擴展 Vertical Extension	指在中日韓統一表意文字基本區及其擴展區增補新表意文字的方式。增補新的表意文字時，必須提供其字源資料。

第二部分 《香港增補字符集—2016》編碼方案

2.1 《香港增補字符集—2016》涵蓋 5 033 個字符，包括經審議通過供電腦平台使用的 4 602 個漢字和 431 個符號。有別於《香港增補字符集—2008》，《香港增補字符集—2016》只提供 ISO/IEC 10646 國際編碼標準碼位，字符表載於附表一。

2.2 《香港增補字符集—2016》包含《香港增補字符集—2008》的 4 579 個漢字和 430 個符號，並增收已納入 ISO/IEC 10646 國際編碼標準字集內的 23 個漢字和 1 個符號，藉此反映香港實際使用漢字的情況。這 23 個漢字已納入中日韓表意文字基本區內，而該符號則納入了貨幣符號區，新增收的字符載於附表二。

2.3 此外，根據 ISO/IEC JTC1/SC2 標準制訂工作組的決議，《香港增補字符集—2008》中有兩個字符的 ISO/IEC 10646 碼位須重新編配如下：

字形 (字源資料)	《香港增補 字符集—2008》	《香港增補 字符集—2016》
鰂 (H-9D73)	4CA4	9FD0
梨 (H-91B5)	3D1D	2A3ED

註：新碼位 9FD0 已納入 2016 年公布的 ISO/IEC 10646:2014 第二修訂版。

4CA4 和 3D1D 兩個碼位會列作兼容碼位，讓尚未採用新版本 ISO/IEC 10646 國際編碼標準的系統能夠繼續使用這些字符。

2.4 下表說明《香港增補字符集》與 ISO/IEC 10646 國際編碼標準各字區之間的關係。表中使用的字區名稱只作說明之用，未必與 ISO/IEC 10646 國際編碼標準文件內所使用的名稱相同。

ISO/IEC 10646 國際編碼標準字區	《香港增補字符集—2008》字符數目	《香港增補字符集—2016》字符數目
符號	430	431
中日韓表意文字基本區	2 291	2 315
擴展區 A	574	572
擴展區 B	1 701	1 702
擴展區 C	1	1
擴展區 D	-	-
擴展區 E	-	-
表意文字兼容字符區	12	12
總數	5 009	5 033

2.5 對於中日韓表意文字兼容字符區的漢字，ISO/IEC 10646 國際編碼標準文件提供與其等同的字符。《香港增補字符集—2016》有 12 個字符收納在 ISO/IEC 10646 表意文字兼容字符區內，與其對應的字符載於附表三，以供參考。

2.6 現有的電腦平台大都支援 ISO/IEC 10646 國際編碼標準。我們建議系統負責人盡快把系統更新至可支援 ISO/IEC 10646 國際編碼標準，以便能更快捷方便地使用最完備的中文字符作通訊和資料交換之用。

第三部分　ISO/IEC 10646 架構下的《香港增補字符集—2016》

3.1　ISO/IEC 10646 國際編碼標準架構下的漢字是指中日韓統一表意文字。中日韓統一表意文字區內每個碼位都有多個字形，並以整列的形式列出，旨在支持和定義某一個國家或地區所需使用的字符，香港使用的漢字載於 H-列。有關 ISO/IEC 10646 國際編碼標準的詳情，可瀏覽以下網址：

http://standards.iso.org/ittf/PubliclyAvailableStandards/ 。

3.2　《香港增補字符集—2016》內所有字符都已納入 ISO/IEC 10646 國際編碼標準之內。ISO/IEC 10646 國際編碼標準會以縱向擴展形式加入新字符，增收該字集並未包含而香港特區有需要使用的字符，也會以橫向擴展形式，為字集內已收錄的字符註明哪些是香港用字，並加入香港字形，藉此反映香港實際使用漢字的情況。

3.3　與《香港增補字符集—2008》比較，《香港增補字符集—2016》增收了 23 個漢字，這些漢字都已納入 ISO/IEC 10646 國際編碼標準，字源資料的編碼格式為“HD-XXXX”(“XXXX”表示該漢字的 ISO/IEC 10646 國際編碼標準碼位)，而香港使用的字形和該字符的字源資料也會加入 H-列之內。

3.4　《香港增補字符集—2016》也增收了 1 個符號，該符號已納入 ISO/IEC 10646 國際編碼標準內，字源資料的編碼格式為“HE-XXXX”(“XXXX”表示該符號的 ISO/IEC 10646 國際編碼標準碼位)。

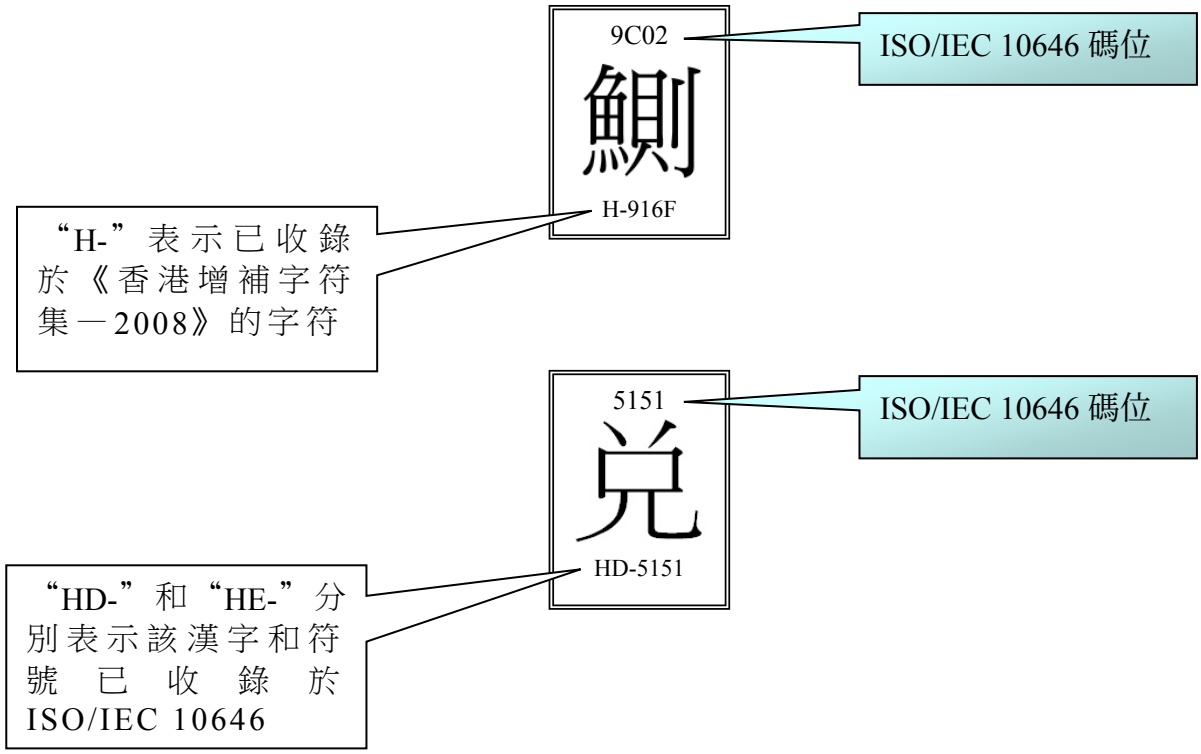
3.5　日後若有新的《香港增補字符集》字符加入 ISO/IEC 10646 國際編碼標準內，該新字的字源資料會以“HC-0001”至“HC-9999”的格式顯示，以表示該字是以縱向擴展形式加入的《香港增補字符集》字符。《香港增補字符集—2016》並沒有以“HC-”定義的字符。

3.6 在 ISO/IEC 10646 國際編碼標準架構下的 H-列，除了列出《香港增補字符集》字符外，也一併列出大五碼字符，字源資料格式為“HB0-XXXX”、“HB1-XXXX”和“HB2-XXXX”，分別代表大五碼符號區、常用字集和次常用字集，當中“XXXX”即該字符的大五碼編碼。有關大五碼編碼對應統一碼的資料，請瀏覽以下網站：http://www.unicode.org 。

附表一：《香港增補字符集—2016》字符表

《香港增補字符集—2016》收錄 5 033 個字符，包括 5 009 個《香港增補字符集—2008》原有的字符，以及新增收的 23 個漢字和 1 個符號。

《香港增補字符集—2016》字符表中一般方格的說明如下：



00A8 **043D**

00A8	00E8	00FC	014D	01D8	0283	0416	0420	042A	0434
¨	è	ü	ō	ǘ	ʃ	Ж	Р	Ъ	д
H-C6D8	H-886F	H-88A2	H-8874	H-887D	H-C8F5	H-C7FA	H-C845	H-C84F	H-C859
00C0	00E9	0100	0153	01DA	028A	0417	0421	042B	0435
À	é	Ā	œ	ǚ	ʊ	З	С	Ы	е
H-8859	H-886D	H-8856	H-C8FA	H-887E	H-C8FD	H-C7FB	H-C846	H-C850	H-C85A
00C1	00EA	0101	016B	01DC	02C6	0418	0422	042C	0436
Á	ê	ā	ū	ǜ	ˆ	И	Т	Ь	ж
H-8857	H-88A7	H-8867	H-8878	H-88A1	H-C6D9	H-C7FC	H-C847	H-C851	H-C85C
00C8	00EC	0112	01CD	0250	0401	0419	0423	042D	0437
È	ì	Ē	Ǎ	ɐ	Ё	Й	У	Э	з
H-885D	H-8873	H-885A	H-8858	H-C8F6	H-C7F9	H-C7FD	H-C848	H-C852	H-C85D
00C9	00ED	0113	01CE	0251	0410	041A	0424	042E	0438
É	í	ē	ǎ	ɑ	А	К	Ф	Ю	и
H-885B	H-8871	H-886C	H-8869	H-886B	H-C7F3	H-C7FE	H-C849	H-C853	H-C85E
00CA	00F2	011A	01D0	0254	0411	041B	0425	042F	0439
Ê	ò	Ě	ǐ	ɔ	Б	Л	Х	Я	й
H-8866	H-8877	H-885C	H-8872	H-C8F8	H-C7F4	H-C840	H-C84A	H-C854	H-C85F
00D2	00F3	011B	01D1	025B	0412	041C	0426	0430	043A
Ò	ó	ě	Ǒ	ɛ	В	М	Ц	а	к
H-8861	H-8875	H-886E	H-8860	H-C8F7	H-C7F5	H-C841	H-C84B	H-C855	H-C860
00D3	00F8	012B	01D2	0261	0413	041D	0427	0431	043B
Ó	ø	ī	ǒ	ɡ	Г	Н	Ч	б	л
H-885F	H-C8FB	H-8870	H-8876	H-88A8	H-C7F6	H-C842	H-C84C	H-C856	H-C861
00E0	00F9	014B	01D4	026A	0414	041E	0428	0432	043C
à	ù	ŋ	ǔ	ɪ	Д	О	Ш	в	м
H-886A	H-887B	H-C8FC	H-887A	H-C8FE	H-C7F7	H-C843	H-C84D	H-C857	H-C862
00E1	00FA	014C	01D6	0275	0415	041F	0429	0433	043D
á	ú	Ō	ǖ	ɵ	Е	П	Щ	г	н
H-8868	H-8879	H-885E	H-887C	H-C8F9	H-C7F8	H-C844	H-C84E	H-C858	H-C863

附表二：《香港增補字符集—2016》新增收的字符

與《香港增補字符集—2008》比較，《香港增補字符集—2016》增收了 23 個漢字和 1 個符號。這些字符以橫向擴展形式納入了 ISO/IEC 10646 國際編碼標準，以反映字符在香港的使用情況，藉此進一步滿足本地中文電子通訊的需要。該 24 個字符表列如下：

序號	香港慣常使用的字形及其 ISO/IEC 10646 碼位	等同的字符及其 ISO/IEC 10646 碼位
1	兑 5151	兌 514C (對應大五碼：0xA749)
2	悦 60A6	悅 6085 (對應大五碼：0xAEAE)
3	挩 635D	捝 6329 (對應大五碼：0xD1BE)
4	敓 655A	敚 6553 (對應大五碼：0xD5A7)
5	梲 68C1	梲 68B2 (對應大五碼：0xD5BF)
6	涗 6D9A	涀 6D97 (對應大五碼：0xD258)
7	税 7A0E	稅 7A05 (對應大五碼：0xB57C)
8	脱 8131	脫 812B (對應大五碼：0xB2E6)
9	蜕 8715	蛻 86FB (對應大五碼：0xB8C0)

序號	香港慣常使用的字形及其 ISO/IEC 10646 碼位	等同的字符及其 ISO/IEC 10646 碼位
10	説 8AAC	說 8AAA (對應大五碼：0xBBA1)
11	鋭 92ED	銳 92B3 (對應大五碼：0xBE55)
12	閱 95B2	閱 95B1 (對應大五碼：0xBE5C)
13	媼 5AAA	媼 5ABC (對應大五碼：0xB6FE)
14	慍 6120	慍 614D (對應大五碼：0xB759)
15	氳 6C32	氳 6C33 (對應大五碼：0xBA72)
16	煴 7174	煴 7185 (對應大五碼：0xE2BE)
17	緼 7DFC	縕 7E15 (對應大五碼：0xEAD5)
18	腽 817D	膃 8183 (對應大五碼：0xE3A6)
19	蕴 85F4	蘊 860A (對應大五碼：0xC4AD)
20	轀 8F3C	轀 8F40 (對應大五碼：0xEEC1)
21	醖 9196	醞 919E (對應大五碼：0xC1DF)

序號	香港慣常使用的字形及其 ISO/IEC 10646 碼位	等同的字符及其 ISO/IEC 10646 碼位
22	告 543F	告 544A (對應大五碼：0xA769)
23	鰊 9C47	---
24	€ 20AC	---

我們在《香港電腦漢字參考字形》的編訂過程中，找到 22 個(序號 1 至 22)與大五碼所取字形不同的香港慣常使用字形，因此在《香港增補字符集—2016》增收了這些漢字。序號 23 的“鰊”字是國際表意文字子集內的字符，“鮫鰊”為食用魚類，是香港常用詞彙，而“鮫”字早已收錄於《香港增補字符集—1999》，因此有需要增收“鰊”字。至於歐元貨幣符號“€”，由於是香港廣泛使用的貨幣符號，故此也一併增收。

附表三：《香港增補字符集—2016》的兼容字符

《香港增補字符集—2016》內有 12 個中日韓兼容字符，下表載列與這些兼容字符對應的字符，以供參考。

序號	《香港增補字符集—2016》內的兼容字符	對應字符
1	龜 F907	龜 9F9C
2	勇 2F825	勇 52C7
3	吆 2F83B	吆 5406
4	罒 2F840	罒 54A2
5	屮 2F878	屮 5C6E
6	彐 2F894	彐 5F22
7	慈 2F8A6	慈 6148
8	晋 2F8CD	晉 6649
9	勞 2F994	芳 82B3
10	夔 2F9B2	夔 456B
11	蝕 2F9BC	蝕 8728
12	貫 2F9D4	貫 8CAB