香港增補字符集－2015

(第六稿)

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

2016 年 6 月

目錄

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 5009 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—2015》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－2015》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。

編碼方案為電子通訊的基礎

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

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，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，特區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 - IRG)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。之前在《香港增補字符集》內已獲編配大五碼碼位的字符則不受影響，可以繼續使用。這項決定旨在進一步向公眾推廣“共通中文界面”，以便利中文電子通訊，也符合“數碼21”資訊科技策略中採用ISO/IEC 10646國際編碼標準的方向。

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。加上在2000年後，電腦平台大多支援ISO/IEC 10646國際編碼標準，《香港增補字符集－2008》遂成為最後一個包含大五碼編碼的《香港增補字符集》版本。

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，中諮會再修訂了《香港增補字符集》字符增收程序和原則，而政府也公布《香港增補字符集－2015》。

《香港增補字符集－2015》的編碼方案和字符表詳載於本文件內。《香港

增字符集－2015》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：http://www.ogcio.gov.hk/tc/business/tech_promotion/ccli/hkscs/。

鳴謝

編訂《香港增補字符集－2015》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”在編訂《香港增補字符集－2015》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－2015》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－2015》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－2015》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－2015》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－2015》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：http://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”。
中日韓表意文字基	ISO/IEC 10646 國際編碼標準中首個為統一表

名詞	定義
本區 CJK Ideographs Main Block	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。
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)。

名詞	定義
政府通用字庫 Government Common Character Set (GCCS)	香港政府於 1995 年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。
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。

名詞	定義
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－2015》編碼方案

2.1 《香港增補字符集－2015》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－2015》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－2015》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－2015》
鰂 (H-9D73)	4CA4	9FD0
黎 (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－2015》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－2015》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－2015》

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－2015》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－2015》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－2015》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－2015》並沒有以“HC-”定義的字符。

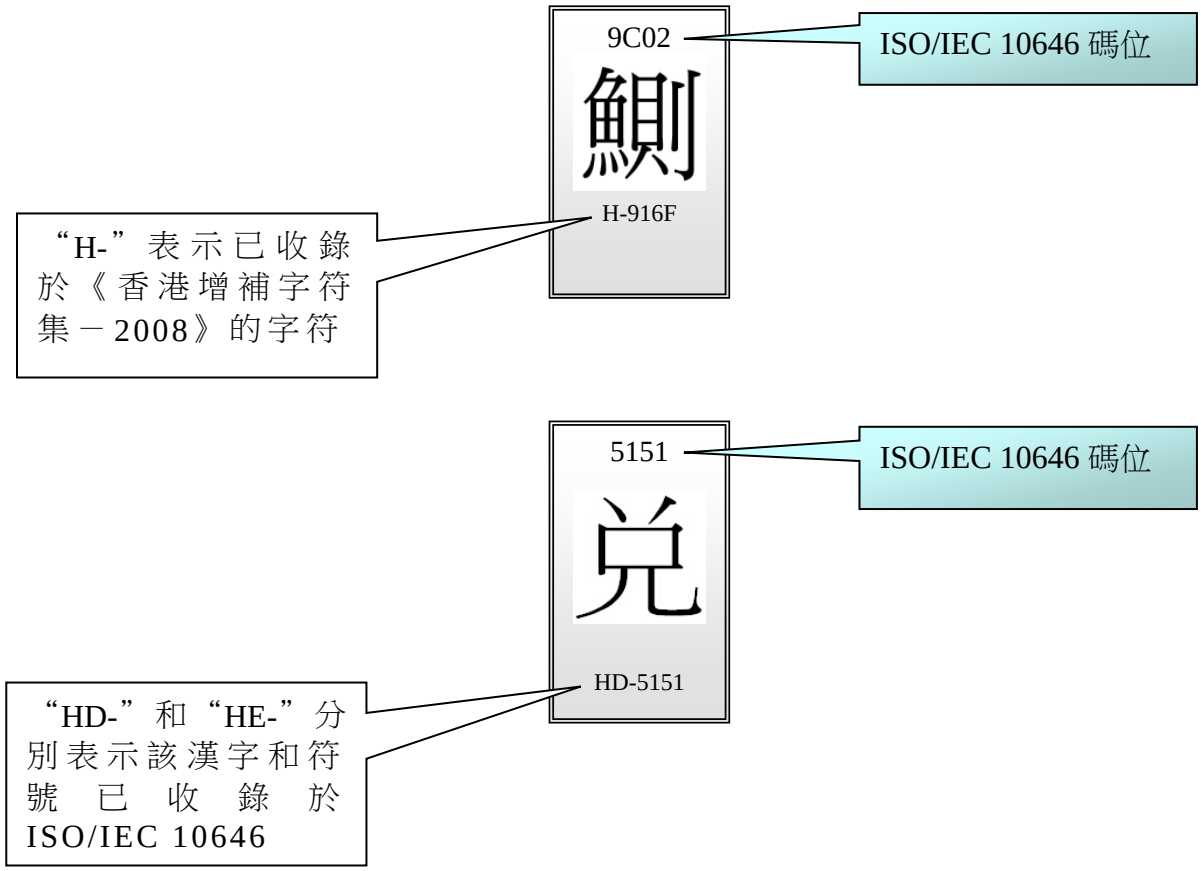
3.6 在 ISO/IEC 10646 國際編碼標準架構下的 H-列，除了列出《香港增補字符集》字符外，也一併列出大五碼字符，字源資料格式為”—“HB0-XXXX”、“HB1-XXXX”和“HB2-XXXX”，分別代表大五碼符號區、常用字集和次常用字集，當中“XXXX”即該字符的大五碼編碼。有關大五碼編碼對應統一碼的資料，請瀏覽以下網站：

http://www.unicode.org。

附表一：《香港增補字符集－2015》字符表

《香港增補字符集－2015》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－2015》字符表中一般方格的說明如下：



附表二: 《香港增補字符集－2015》新增收的字符

與《香港增補字符集－2008》比較，《香港增補字符集－2015》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	棁 68B2 (對應大五碼：0xD5BF)
6	涗 6D9A	涚 6D97 (對應大五碼：0xD258)
7	稅 7A0E	稅 7A05 (對應大五碼：0xB57C)
8	脫 8131	脫 812B (對應大五碼：0xB2E6)
9	蛻 8715	蛻 86FB (對應大五碼：0xB8C0)
10	誽 8AAC	說 8AAA (對應大五碼：0xBBA1)
11	鋭 92ED	銳 92B3 (對應大五碼：0xBE55)

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
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	縕 7DFC	縕 7E15 (對應大五碼：0xEAD5)
18	膃 817D	膃 8183 (對應大五碼：0xE3A6)
19	蘊 85F4	蘊 860A (對應大五碼：0xC4AD)
20	轀 8F3C	轀 8F40 (對應大五碼：0xEEC1)
21	醞 9196	醞 919E (對應大五碼：0xC1DF)
22	告 543F	告 544A (對應大五碼：0xA769)
23	鰼 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－2015》增收了這些漢字。序號 23 的“鰼”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－2015》的兼容字符

《香港增補字符集－2015》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－2015》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	中 5C6E
6	戓 2F894	戓 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

Hong Kong Supplementary Character Set - 2015

(Draft 6)

Office of the Government Chief Information Officer &
Official Languages Division, Civil Service Bureau
The Government of the Hong Kong Special Administrative Region

June 2016

Table of Contents

Preface

After the first release of the Hong Kong Supplementary Character Set (HKSCS) in 1999, there have been three updated versions. The HKSCS-2001, HKSCS-2004 and HKSCS-2008 were published with 116, 123 and 68 new characters added respectively. A total of 5 009 characters were included in the HKSCS-2008. These publications formed the foundation for promoting the adoption of the ISO/IEC 10646 international coding standard, and were widely supported and adopted by the IT sector and members of the public.

The ISO/IEC 10646 international coding standard is developed by the International Organization for Standardization (ISO) to provide a common technical basis for the storage and exchange of electronic information. It provides a unified standard for the coding of characters in all major languages in the world including traditional and simplified Chinese characters. Containing more than 80 000 Han characters, the ISO/IEC 10646:2014 provides computer platforms with comprehensive support. However, supporting a character set with over 80 000 Han characters will unnecessarily increase the cost and time of product development. Hence, vendors of font or input method software will select an appropriate number of characters in the light of the requirements of individual countries and regions in developing supporting products.

All the characters in the HKSCS-2015 have already been included in the ISO/IEC 10646 international coding standard. ISO/IEC 10646 will continue adding new characters that have not been included in its character set but are used in Hong Kong (i.e. the addition of new characters through vertical extension). Furthermore, for characters already included in the ISO/IEC 10646, information about the characters and glyphs commonly used in Hong Kong will be added to reflect the actual use of these characters locally (i.e. the addition of information to existing characters through horizontal extension). This will not only facilitate the development of vendor support for Chinese characters actually used in Hong Kong and the relevant localised technology, but will also reduce the time and cost of development, enabling the IT industry to develop more products suitable for Hong Kong.

Compared with the HKSCS-2008, the HKSCS-2015 has 23 more Chinese characters and one more symbol, further fulfilling the needs of local electronic communication in Chinese.

Coding Scheme as the Foundation of Electronic Communication

Information stored in a computer or transmitted in electronic communication is coded according to a pre-defined coding scheme. For information in Chinese, as early as the 1990s, there were different coding schemes including Big-5, GB (Guo Biao) and the ISO/IEC 10646 international coding standard. At that time, as the coding schemes did not cover all the Chinese characters commonly used in Hong Kong, users might need to create unencoded characters on their computers and assign internal codes for them in the user-defined area. Some characters were used in the names of persons and places while some in the Cantonese dialect. This worked well in stand-alone computers, but when computers were connected to each other, such user-defined characters might give rise to problems in communication and data exchange.

Common Chinese Language Interface

Under the Digital 21 Strategy for IT Development, the Government has developed an open and common Chinese language interface for the Hong Kong Special Administrative Region (HKSAR), with the objective of achieving more accurate electronic communication in Chinese. A pivotal element of the open and common Chinese language interface is the adoption of the ISO/IEC 10646 international coding standard.

Development of the HKSCS

To facilitate electronic communication within the Government, the Hong Kong Government developed the Government Common Character Set (GCCS) in 1995. The release of the GCCS marked the first step in coordinating the adoption of user-defined Chinese characters and it was well received by the public as a supplement to the standard character set of Big-5. This common character set was later enhanced by the Government in collaboration with the Chinese Language Interface Advisory Committee (CLIAC), which comprises representatives from academia, language and linguistics associations, the information technology industry and the publishing industry. The enhanced character set included characters collected from various sectors in the HKSAR and represented a common set for the community. It was named the Hong Kong Supplementary Character Set (HKSCS) and was published in September 1999. This version had 4 702 characters and was also known as the

HKSCS-1999 for aligning with the versions published afterwards.

The Government recognised the need for the public and government departments to include new characters in the HKSCS from time to time. In collaboration with CLIAC, the Government published the procedures and principles for the inclusion of characters in the HKSCS in April 2000. CLIAC meets regularly to consider applications for inclusion of characters in the HKSCS. Once approved, the new HKSCS characters will be submitted to the Ideographic Rapporteur Group, a working group under the ISO, for inclusion in the future releases of the ISO/IEC 10646 standard.

The HKSCS has subsequently been updated three times and a total of 5 009 characters were included in the HKSCS-2008. There were two code allocation schemes for each of the four versions of the HKSCS: one for the Big-5 which was used in Hong Kong at that time and the other for ISO/IEC 10646. This arrangement aimed at a gradual migration of computing platforms from Big-5 to ISO/IEC 10646.

With the emergence of more platforms and products supporting ISO/IEC 10646, the Government promulgated the revised procedures and principles for the inclusion of characters in the HKSCS in April 2008. Since then, for newly included HKSCS characters, only code points of the characters in ISO/IEC 10646 will be provided and no Big-5 code points will be assigned. All HKSCS characters already assigned Big-5 code points will not be affected. This decision aimed to further promote the Common Chinese Language Interface, so as to facilitate electronic communication in Chinese. This is also in line with the Digital 21 Strategy of adopting the ISO/IEC 10646 international coding standard.

The ISO/IEC 10646:2011 has included all characters in the HKSCS-2008, indicating a complete integration of HKSCS-2008 with ISO/IEC 10646 and marked a milestone for the full adoption of the ISO/IEC 10646. Moreover, with computer platforms generally supporting ISO/IEC 10646 after 2000, the HKSCS-2008 became the last HKSCS version with Big-5 code points.

In order to reflect the local use of Han characters in the ISO/IEC 10646, the Government published the HKSCS-2015 upon further revisions to the procedures and principles for the inclusion of characters in the HKSCS were made by CLIAC.

The coding scheme and code table of the HKSCS-2015 are provided in this document. Other mapping tables of the HKSCS-2015, and the documents stipulating the procedures and principles for the inclusion of characters in the HKSCS can be found at the Common Chinese Language Interface website at:
http://www.ogcio.gov.hk/en/business/tech_promotion/ccli/hkscs/.

Acknowledgement

This specification was completed with invaluable assistance from CLIAC. We would like to thank the following committee members for their contribution: current members Ms. Michelle CHEUNG, Mr. CHEUNG To, Mr. Ricky CHUN, Dr. FUNG Suk Yee Roxana, Mr. Alan KAN, Dr. LEE Ka Kui, Miss LEUNG Wai Chu Judy, Professor LU Qin, M.H., Dr. MAN Ying Ha, Dr. TANG Pui Ling, Mr. Wilson WONG, and former committee members: Mr. CHIU Leung Fai Fritz, Dr. FONG Kin Kiu Ken, Mr. Raymond HUNG, Mr. LEUNG Shung Yu and Mr. Jonathan SHEA.

We would also like to thank the working group chaired by Professor LU Qin for their great efforts. The working group consists of some CLIAC members and Dr. CHEUNG Kwan Hin, Mr. LAI Tat Kiu Alex, Mr. LEE Kin Hong, Dr. LUN Suen Caesar, Professor WONG Yiu Kwan and representative from the Education Bureau.

Section 1 Overview

1.1 This document provides the characters in the HKSCS-2015 and their corresponding code points in ISO/IEC 10646, and explains the overall coding architecture of the HKSCS-2015 in the ISO/IEC 10646 international coding standard. The HKSCS-2015 is fully compatible with the GCCS and the previous versions of the HKSCS.

1.2 The HKSCS-2015 contains 5 033 characters, including 5 009 characters from the HKSCS-2008, and the newly added 23 Chinese characters and one symbol.

1.3 The HKSCS-2015 is a coded character set. It is not meant to be a glyph standard. For glyph guidelines, please consult the "Reference Glyphs for Chinese Computer Systems in Hong Kong", which is available at: http://www.ogcio.gov.hk/en/business/tech_promotion/ccli/download_area/.

1.4 For the purpose of this document, the following definitions of terms will apply:

Term	Definition
Basic Multilingual Plane (BMP, Plane 0)	The first code plane in the ISO/IEC 10646 coding framework (i.e. "Plane 0" or basic plane). Code points are from 0000 to FFFF.
Block	A collection of characters that share common characteristics.
Character	A member of a set of elements used for the organisation, control or representation of data.
Character Glyph	In ISO/IEC 10646, it refers to a Han character in its abstract form as an image. It is independent of any specific image. The basic elements to form an ideograph are strokes, radicals, components and their relative positions.

Term	Definition
Character Set	A defined set of characters.
CJK Compatibility Ideographs	An area defined in the BMP (Plane 0) for compatibility with CJK Ideographs Blocks. This area is used to include the variants or duplicate characters already coded in CJK Ideograph Sources which would otherwise not be coded in ISO/IEC 10646. Code points are from F900 to FAFF. In ISO/IEC 10646, these variants and their corresponding standard characters are unified. However, they are assigned different code points in their respective CJK Ideograph Sources already. Therefore, this special area is defined to avoid having one character with multiple code points in CJK Ideographs Blocks and at the same time allow round-trip conversion for backward compatibility. Every Compatibility Ideograph has a corresponding standard character coded in CJK Ideographs Blocks.
CJK Compatibility Ideographs Supplement	An extended area defined in the Supplementary Ideographic Plane (SIP, Plane 2) for compatibility with CJK Ideographs Blocks. Code points are from 2F800 to 2FA1F. The Han characters in the compatibility blocks and Ideographs Supplement are collectively referred to as “compatibility characters” in this document.
CJK Ideographs Main Block	The first block assigned to the unified ideographs, including Chinese, Japanese and Korean characters. Code points are from 4E00 to 9FFF.
CJK Ideograph Source	The CJK ideographs in the ISO/IEC 10646 international coding standard are defined based on the original computer character standards of China, Japan, Korea and other countries and regions. The original computer character

Term	Definition
	standard or specification is called CJK Ideograph Source. The countries and regions are represented by letters as follows: Mainland China (G), Hong Kong (H), Japan (J), South Korea (K), Singapore (S), Taiwan (T) and Vietnam (V).
CJK Unified Ideographs	A set of ideographs defined in the ISO/IEC 10646 international coding standard in accordance with the unification rules. The ideographs are derived from the original character standards of China, Japan, Korea, and other countries and regions. As the first version of the standard includes ideographs mainly from China, Japan and Korea, the name “CJK” has been used ever since. In this document, these ideographs are also referred to as “Han characters”.
Code Point	An assigned hexadecimal code value to represent a character.
Coded Character Set	A character set established under a set of unambiguous rules. It defines the relationship between the characters of the set and their coded representation.
CJK Ideographs Extension Blocks	The blocks developed as extensions to the CJK Ideographs Main Block. Extension A Block is placed on the BMP and the subsequent extension blocks are on the Supplementary Ideographic Plane (SIP, Plane 2).
Government Common Character Set (GCCS)	A coded character set developed by the Hong Kong Government in 1995 for exchanging and processing Chinese information within the Government.

Term	Definition
Horizontal Extension	This refers to the addition of information and source reference to the characters already included in the ISO/IEC 10646.
H-Column	Each code point of the CJK Unified ideographs has multiple glyphs and these glyphs are listed in individual columns. This multi-column format aims to support and define the characters used in a particular country or region. The Chinese characters used in Hong Kong are listed in the H-column.
Ideograph	Refers to a character in a writing system in which the scripts are not primarily used to represent sound, but to represent meaning. Chinese characters are ideographs.
ISO/IEC 10646	An ISO standard on a coded character set. It aims at providing one single character set to encompass the characters of all major languages.
Source Reference	A reference established by associating a CJK Ideograph code point with one or several values in the source standards from which the CJK Unified Ideographs in ISO/IEC 10646 are derived.
Supplementary Ideographic Plane (SIP, Plane 2)	Plane 2 is assigned under the ISO/IEC 10646 coding framework for CJK ideograph extensions. Code points are from 20000 to 2FFFF.
Unification	The process of assigning one code point to two or more character glyphs which, though seemingly different, are actually variants representing the same element in data representation. Consequently, only one of the variants is

Term	Definition
	selected as the representative.
Vertical Extension	A method for adding new ideographs to the CJK Ideographs Main Block and other extension blocks. Source references are required when new ideographs are added.

Section 2 Coding Scheme of the HKSCS-2015

2.1 The HKSCS-2015 consists of 5 033 characters, including 4 602 Chinese characters and 431 symbols. All these characters have been reviewed for use on computer platforms. Unlike the HKSCS-2008, the HKSCS-2015 provides code points of the characters in the ISO/IEC 10646 only. The code table can be found at Table 1.

2.2 The HKSCS-2015 contains all the 4 579 Chinese characters and 430 symbols from the HKSCS-2008, and the newly added 23 Chinese characters and one symbol already included in the ISO/IEC 10646, so as to reflect the actual use of these characters in the HKSAR. These 23 Chinese characters are included in the CJK Ideographs Main Block and the symbol is in the block for currency symbols. The newly included characters are listed in Table 2.

2.3 According to the resolution made by the working group under the ISO/IEC JTC1/SC2, two characters from the HKSCS-2008 should have their code points in the ISO/IEC 10646 re-assigned as follows:

Glyph (source reference)	HKSCS-2008	HKSCS-2015
鰂 (H-9D73)	4CA4	9FD0
黎 (H-91B5)	3D1D	2A3ED

The new code point 9FD0 has been adopted in the ISO/IEC 10646:2014/Amendment 2:2016.

Code points 4CA4 and 3D1D are kept as compatibility points to enable computer systems yet to adopt the latest version of the ISO/IEC 10646 to continue using them.

2.4 The table below shows the relationship between the HKSCS and the character blocks of the ISO/IEC 10646 coding standard. The names of the blocks given are for ease of reference only and may not be the same as those used in the ISO/IEC 10646 international coding standard document.

ISO/IEC 10646 Character block	Number of characters in the HKSCS-2008	Number of characters in the HKSCS-2015
Symbols	430	431
CJK Ideographs Main Block	2 291	2 315
Extension A	574	572
Extension B	1 701	1 702
Extension C	1	1
Extension D	-	-
Extension E	-	-
Compatibility Ideograph Block	12	12
Total	5 009	5 033

2.5 The ISO/IEC 10646 document provides compatible characters for characters included in the CJK Compatibility Ideograph Block, in which 12 are HKSCS-2015 characters. These 12 characters and their corresponding characters are listed in Table 3 for reference.

2.6 As most of the existing computer platforms support ISO/IEC 10646, persons-in-charge are recommended to upgrade their systems to support ISO/IEC 10646 as soon as possible to enable more efficient and convenient use of the most comprehensive Chinese character set for communication and information exchange.

Section 3 HKSCS-2015 under the Architecture of the ISO/IEC 10646

3.1 Under the architecture of the ISO/IEC 10646 international coding standard, Han characters refer to the CJK unified ideographs. Each code point of the CJK Unified ideographs has multiple glyphs listed in individual columns. This multi-column format serves to support and define the characters needed in a particular country or region. The Chinese characters used in Hong Kong are listed in the H-column. Details of the ISO/IEC 10646 international coding standard are available at: http://standards.iso.org/ittf/PubliclyAvailableStandards/.

3.2 All the HKSCS-2015 characters have been included in the ISO/IEC 10646 international coding standard. Characters used in Hong Kong but not included in the standard will be added through vertical extension. For characters already in the standard, information will be added to specify which characters are used in Hong Kong and the glyphs used locally will be included through horizontal extension, so as to reflect the actual use of Chinese characters in Hong Kong.

3.3 Based on the HKSCS-2008, the HKSCS-2015 has newly included 23 Chinese characters already in the ISO/IEC 10646. These characters are denoted as "HD-XXXX" (where "XXXX" is the code point of the character in ISO/IEC 10646). The preferred glyphs used in Hong Kong and the source reference are also added to the H-column.

3.4 The HKSCS-2015 has also added one new symbol, which is already included in the ISO/IEC 10646. This symbol is denoted in the form of "HE-XXXX", where "XXXX" is the code point of the symbol in the ISO/IEC 10646.

3.5 For any HKSCS character to be added to the ISO/IEC 10646 in the future, the source reference will be given in the form of HC-0001 to HC-9999, denoting that the character is included in the ISO/IEC 10646 through vertical extension. The HKSCS-2015 does not include characters with source reference denoted in the form of "HC-".

3.6 Under the architecture of the ISO/IEC 10646, the H-column lists not only the characters in the HKSCS, but also those included in the Big-5 coding scheme. The source reference of such Big-5 characters ~~are~~ is provided in the form of "HB0-XXXX", "HB1-XXXX" and "HB2-~~–~~XXXX ", where "XXXX" is the code point of the character under the Big-5 coding scheme, denoting characters from the Big-5 symbol area, frequently used characters and less frequently used characters respectively. Information on the mapping of the Big-5 and Unicode is available at the following website:
http://www.unicode.org.

Table 1: Code Table of the HKSCS-2015

The HKSCS-2015 contains 5 033 characters, including 5 009 characters from the HKSCS-2008, and 24 newly added characters (23 Chinese characters and one symbol).

The following are examples of typical cells in the code table of the HKSCS-2015:

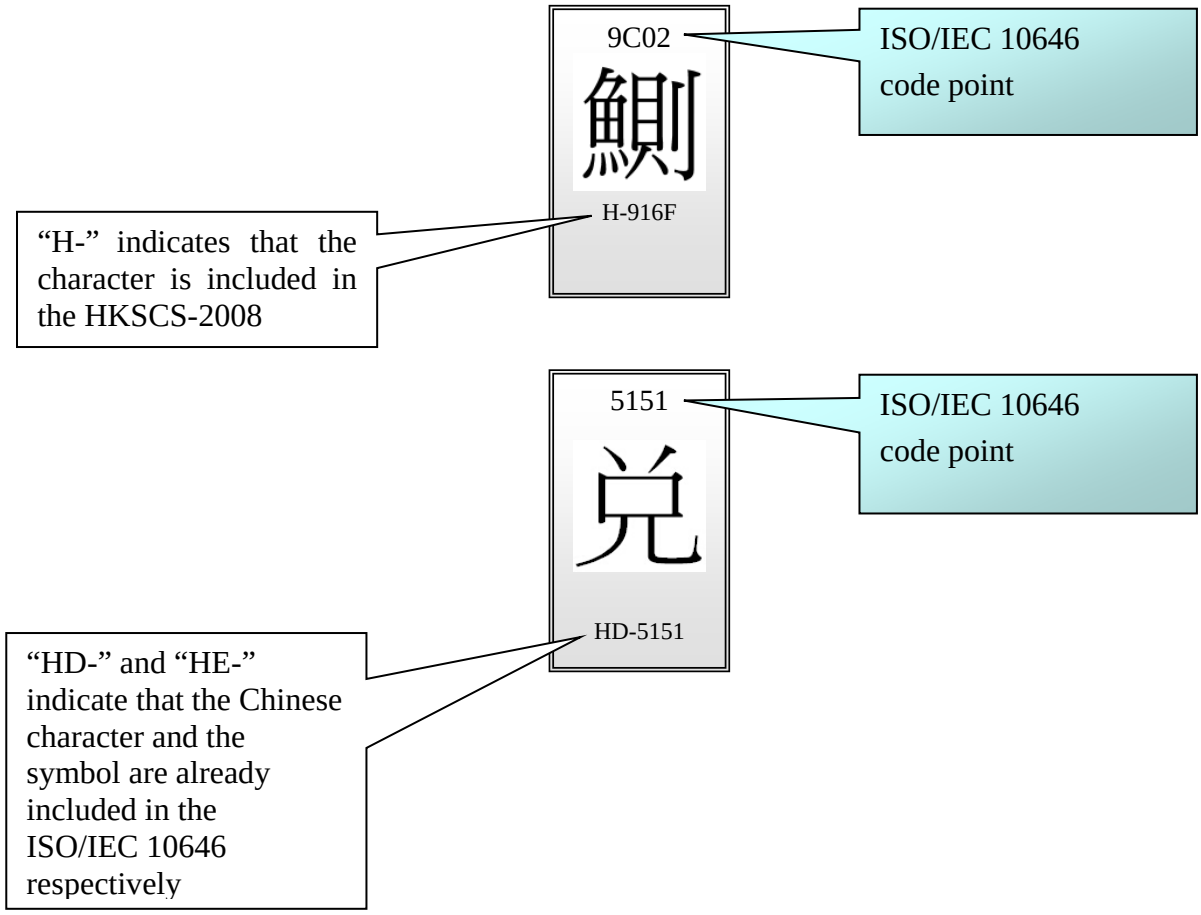


Table 2: Newly Included Characters in the HKSCS-2015

Compared with the HKSCS-2008, the HKSCS-2015 includes 23 more Chinese characters and one more symbol. These characters are included in the ISO/IEC 10646 through horizontal extension, reflecting their use in Hong Kong and further fulfilling the needs of the local electronic communication in Chinese. These 24 characters are listed below.

S/N	Glyph commonly used in Hong Kong and its ISO/IEC 10646 code point	Unifiable character and its ISO/IEC 10646 code point
1	兌 5151	兌 514C (corresponding Big-5 code: 0xA749)
2	悦 60A6	悅 6085 (corresponding Big-5 code: 0xAEAE)
3	挩 635D	捝 6329 (corresponding Big-5 code: 0xD1BE)
4	敓 655A	敚 6553 (corresponding Big-5 code: 0xD5A7)
5	梲 68C1	棁 68B2 (corresponding Big-5 code: 0xD5BF)
6	涗 6D9A	涚 6D97 (corresponding Big-5 code: 0xD258)
7	稅 7A0E	稅 7A05 (corresponding Big-5 code: 0xB57C)
8	脫 8131	脫 812B (corresponding Big-5 code: 0xB2E6)
9	蛻 8715	蛻 86FB (corresponding Big-5 code: 0xB8C0)
10	誽 8AAC	說 8AAA (corresponding Big-5 code: 0xBBA1)
11	鋭 92ED	銳 92B3 (corresponding Big-5 code: 0xBE55)

S/N	Glyph commonly used in Hong Kong and its ISO/IEC 10646 code point	Unifiable character and its ISO/IEC 10646 code point
12	閱 95B2	閱 95B1 (corresponding Big-5 code: 0xBE5C)
13	媼 5AAA	媼 5ABC (corresponding Big-5 code: 0xB6FE)
14	慍 6120	慍 614D (corresponding Big-5 code: 0xB759)
15	氳 6C32	氳 6C33 (corresponding Big-5 code: 0xBA72)
16	煴 7174	煴 7185 (corresponding Big-5 code: 0xE2BE)
17	緼 7DFC	縕 7E15 (corresponding Big-5 code: 0xEAD5)
18	腽 817D	膃 8183 (corresponding Big-5 code: 0xE3A6)
19	蘊 85F4	蘊 860A (corresponding Big-5 code: 0xC4AD)
20	轀 8F3C	轀 8F40 (corresponding Big-5 code: 0xEEC1)
21	醞 9196	醞 919E (corresponding Big-5 code: 0xC1DF)
22	告 543F	告 544A (corresponding Big-5 code: 0xA769)
23	鱇 9C47	---
24	€ 20AC	---

During the formulation of the Reference Glyphs for Chinese Computer Systems in Hong Kong, it is found that for some characters (characters S/Ns 1 to 22), there are differences between the glyphs commonly used in Hong Kong and those specified in the Big-5 code table. These glyphs used in Hong Kong are therefore included in the HKSCS-2015. As “鱇” (character S/N 23) is included in the International Ideographs Core and commonly used in

Hong Kong together with “鮫”, a HKSCS-1999 character, to form the name of the food fish “鮫鰊”, and as the euro sign “€” is a commonly used currency symbol in Hong Kong, they are also included in the HKSCS-2015.

Table 3: Compatibility Characters in the HKSCS-2015

The HKSCS-2015 contains 12 characters which are in the CJK Compatibility Ideographs Block. Their corresponding characters are shown in the following table for reference.

S/N	Compatibility characters in the HKSCS-2015	Corresponding character
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	蟲 2F9BC	蟲 8728
12	貫 2F9D4	貫 8CAB