

香港增補字符集－2015

(第五稿)

2016年XX月

目錄

序言

第一部分 概論 1 - 1

第二部分 《香港增補字符集－2015》編碼方案..... 2 - 1

第三部分 ISO/IEC 10646 架構下的《香港增補字符集－2015》 3 - 1

附表一: 《香港增補字符集－2015》字符表 I - 1

附表二: 橫向擴展的《香港增補字符集－2015》字符..... II - 1

附表三: 《香港增補字符集－2015》的兼容字符 III - 1

序言

《香港增補字符集》自 1999 年初版，其後三度更新，出版《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，為協調中文字符方面和過渡致 ISO/IEC 10646 國際編碼標準的資訊科技策略定下基礎，亦得到資訊科技業界和市民的支持和使用。

ISO/IEC 10646 國際編碼標準是由國際標準化組織 (International Organisation for Standardisation, ISO) 發展的一套應用於儲存及交換電子資料的國際編碼標準，涵蓋各種主要語文的字符，包括繁體及簡體的中文。ISO/IEC 10646:2014 國際編碼標準已包含漢字逾 80 000 個，為電腦平台提供整體而全面的支援。然而支持全部八萬多字的字集，不僅會增加開發時間和費用，在一般情況下，實際應用上亦非必要。所以字體或輸入法開發商會因應個別國家地區的用字要求，選取適量字符開發支援產品。

《香港增補字符集—2015》的出版，除了繼續在 ISO/IEC 10646 增收該標準並未包含而香港有需要使用的字符外(對於 ISO/IEC 10646 國際編碼標準而言，加入新字符是一種縱向的擴展)，還會在 ISO/IEC 10646 國際編碼標準字集內已收納的字符註明哪些字符是香港用字及加入香港字形，藉此反映香港實際使用漢字的情況(對於 ISO/IEC 10646 國際編碼標準而言，只新增了該字符是香港用字的資訊及字形，而並不是新增字符，可視為對 ISO/IEC 10646 一種橫向的擴展)。這不但方便廠商支援香港實際使用的漢字及發展相應的本地化技術，更可以減少開發成本和時間，有利資訊行業開發多更適合香港使用的產品。

相比《香港增補字符集—2008》，《香港增補字符集—2015》主要增收了 22 個橫向擴展的漢字[和 XX 個符號]，藉此進一步完善本地中文電子通訊的需要。

編碼方案為電子通訊的基礎

將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把香港需要使用的所有中文字符包含在內，當電腦用戶需要在電腦上使用這些字符時，唯有自行在使用者造字區為這些字符編配碼位。這些字符之中，很多都是香港特區常用的，包括人名、地名和粵方言用字。這種做法，對獨立運作的電腦不會構成問題，但假如這些電腦互相連接起來，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用的中文界面，令中文電子通訊能更加準確。共通中文界面的重點是採用 ISO/IEC 10646 國際編碼標準。ISO/IEC 10646 國際編碼標準由國際標準化組織統籌制訂，目的是建立一套涵蓋世界主要語文的字符集，包括繁簡體中文字。

《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，取名為《香港增補字符集》（為了與之後發表

的擴充版一致，也稱《香港增補字符集－1999》），共收錄 4702 個字符，成為社會人士共用的字符集。

2001 年 12 月，政府公布《香港增補字符集－2001》。這個版本較《香港增補字符集－1999》增收了 116 個字符。2005 年 5 月，政府公布《香港增補字符集－2004》，增收了 123 個字符，2008 年 4 月，政府公布《香港增補字符集－2008》，增收了 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。

這四版的《香港增補字符集》各版本有兩套編碼方案，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。這是為逐漸由大五碼系統過渡至 ISO/IEC 10646 系統的安排。2000 年後，由於 ISO/IEC 10646 電腦系統已逐漸普及，所以《香港增補字符集－2008》是《香港增補字符集》包括大五碼編碼的最後版本。

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。

《香港增補字符集》字符增收程序和原則的修訂

鑑於市民和政府部門不時需要把新字符納入《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 國際編碼標準日後的版本中，對於 ISO/IEC 10646 國際編碼標準而言，加入新字符是一種縱向擴展。

隨着支援 ISO/IEC 10646 的平台及產品普及，2008 年 4 月政府公布修訂

《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 國際編碼標準碼位，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，一概不受影響，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。

為了在 ISO/IEC 10646 更能反映香港實際使用的漢字情況的需要，中諮會修訂了《香港增補字符集》字符增收程序和原則，和出版《香港增補字符集－2015》。

《香港增補字符集－2015》的編碼方案和字符列表詳載於本文件內。《香港增補字符集－2015》相關的對應表，《香港增補字符集》字符增收程序和《香港增補字符集》字符增收原則等文件可在“共通中文界面”網站下載：

http://www.ogcio.gov.hk/tc/business/tech_promotion/ccli/hkscs/。

鳴謝

待補充。

第一部分 概論

- 1.1 本文件旨在說明《香港增補字符集－2015》在 ISO/IEC 10646 中的整體編碼架構，並提供字符集在 ISO/IEC 10646 的對應碼位。《香港增補字符集－2015》與各版本《香港增補字符集》和《政府通用字庫》完全兼容。
- 1.2 《香港增補字符集－2015》包含 5 031 個字符，其中 5 009 個是《香港增補字符集－2008》原有的字符，新增收的字符共有 22 個 [和 XX 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。本文件所列出的漢字字形則遵從《香港電腦漢字參考字形》。該《香港電腦漢字參考字形》文件可於以下網址下載：
http://www.ogcio.gov.hk/tc/business/tech_promotion/ccli/download_area/。
- 1.4 以下是在本文件中出現的一些名詞及相應的定義：

名詞	定義
基本多文種平面 (平面 0) Basic Multilingual Plane (BMP, Plane 0)	ISO/IEC 10646 編碼架構內的基本多文種平面，簡稱基本平面，也稱“平面 0”，碼位由 0000 至 FFFF。
字區 Block	一個序列的編碼區域，所對應的字符具有共同特徵。
字符 Character	用於組織、控制或表示數據的諸元素集合中的一員。

名詞	定義
字形 Character Glyph	在 ISO/IEC 10646 中特指抽象可識別的圖形符號，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、漢字部首與部件以及它們的位置關係等。
字符集 Character Set	一組已定義的字符。
中日韓兼容區表意文字 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及其對應的標準字符認同、但在收錄時，其原字集內已獲編配兩個碼位。因此 ISO/IEC 10646 建立了特殊的編碼兼容區用以保證在 ISO/IEC 10646 中，中日韓表意文字區不出現一個字符多個碼位，同時兼顧原字集和 ISO/IEC 10646 內雙向轉換。每個兼容區內的字符都有其對應的中日韓表意文字區內的編碼和字符。
中日韓兼容擴展區表意文字 CJK Compatibility Ideographs Supplement	該區為中日韓兼容區的擴展區，位於增補表意文字平面（平面 2），碼位由 2F800 至 2FA1F。本文件中將兼容區和兼容擴展區的漢字簡稱為“兼容漢字”。
中日韓表意文字基本區	在 ISO/IEC 10646 平面 0 中最初定義的認同的中日韓表意文字區，碼位由 4E00 至 9FFF。

名詞	定義
CJK Ideographs Main Block	
原字集 CJK Ideograph Source	ISO/IEC 10646 國際編碼標準內的中日韓文字是在整合中國、日本、韓國及其他國家和地區原有的電腦編碼標準內的文字後得出的。這些原有的電腦編碼標準或規範內的文字稱為原字集。原字集的頒布國家和地區均有代號：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	ISO/IEC 10646 國際編碼標準包括的中國、日本、韓國及其他國家和地區經認同規則整合後所採納的漢字字符的統稱。最早的表意文字來源主要是中國、日本和韓國，因此“中日韓”名稱沿用至今。本文也簡稱為“漢字”。
碼位 Code Point	代表一個字符的二進制碼。由於其他字符集的編碼也會以二進制碼代表，如有須要本文會在 ISO/IEC 10646 的碼位用一個前置的“U+”表示，藉以識別。
編碼字符集 Coded Character Set	一套無歧義的規則，用以建立一個字符集，並確定該字符集中每一個字符與其對應的碼位關係。
中日韓表意文字擴展區 CJK Ideographs Extension	指中日韓表意文字基本區之後逐漸擴充的中日韓表意文字字符區，擴展區 A 置放於基本多文種平面（平面 0），其他的擴展區置放於增補表意

名詞	定義
Blocks	文字平面（平面 2）。
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	橫向擴展用於在 ISO/IEC 10646 已收納的字符上增補某一國家或地區所使用的字形和字源資料。
H-列 H-Column	中日韓統一表意文字區內每一個碼位以多字並列的形式來支持和定義某一個國家或地區所需的漢字。香港使用的漢字羅列於 H-列。
表意文字 Ideograph	指書寫系統主要是表其“意”而不表其“聲”，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字符集。建立這套標準的目的，是編訂包含所有主要語言的單一字符集。
ISO/IEC 10646:2003	指於 2004 年發表的 ISO/IEC 10646 國際編碼標準。該版本合併了 ISO/IEC 10646-1:2000 及其補充版本 ISO/IEC 10646-2:2001 而成為單一的發行本。
私人使用區 Private Use Area (PUA)	在 ISO/IEC 10646 國際編碼標準中預留給用戶

名詞	定義
	和廠商自行定義字符的字區。原則上，私人使用區只供個別用戶和廠商作私人用途，不用作資料交換。
字源資料 Source Reference	ISO/IEC 10646 中的中日韓表意文字是從多個不同的國家和地區提出的提案中整理出來的字符，這些所使用的國家和地區的來源索引即字源資料。
增補表意文字平面 (平面 2) Supplementary Ideographic Plane (SIP, Plane 2)	ISO/IEC 10646 編碼架構中將“平面 2”用作表意文字的增收，碼位由 20000 至 2FFFF。
認同 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。各有關的符號代表同一數據元素，因此只選擇其中一個作為代表。
縱向擴展 Vertical Extension	縱向擴展用於增補新的表意文字至中日韓統一表意文字區及其擴展區，增補新的表意文字時，需要提供其字源資料。

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》涵蓋香港經審議通過供電腦平台使用的 4601 個漢字和 430 個符號，總共包含 5031 個字符。《香港增補字符集－2015》只有提供 ISO/IEC 10646 碼位，字符列表見於附表一。
- 2.2 《香港增補字符集－2015》包含了來自《香港增補字符集－2008》的 430 個符號和 4579 個漢字，並增收了 22 個 ISO/IEC 10646 內已有的漢字，註明哪些字符是香港用字，藉此反映香港實際使用的漢字。這 22 個字符全都在中日韓表意文字基本區內，字符列表見於附表二。
- 2.3 根據 ISO/IEC JTC1/SC2 標準制訂工作組的決議，對兩個《香港增補字符集－2008》字符所編配的 ISO/IEC 10646 碼位進行如下的修改：

字形 (字源資料)	《香港增補字符集－ 2008》	《香港增補字符集－ 2015》
𩶛 (H-9D73)	4CA4	9FD0#
梨 (H-91B5)	3D1D	2A3ED

#新碼位 9FD0 已獲通過，待在新版本的 ISO/IEC 10646 正式發表

《香港增補字符集－2015》保留 4CA4 和 3D1D 作相應的兼容碼位，以方便未更新採用新版本 ISO/IEC 10646 的系統使用這些字符。

- 2.4 就中日韓兼容表意文字，ISO/IEC 10646 標準文件提供可與之認同的字符。《香港增補字符集－2015》有 12 個中日韓兼容表意文字，與之認同的字符列在附表三以供參考。
- 2.5 下表說明《香港增補字符集》與 ISO/IEC 10646 的字區之間的關係。表中使用的字區名稱只作說明之用，與 ISO/IEC 10646 標準文件內所使用的名稱不一定相同。

ISO/IEC 10646 字 區	《香港增補字符集—2008》	《香港增補字符集—2015》
符號	430	430
中日韓表意文字基本區	2291	2314
擴展區 A	574	572
擴展區 B	1701	1702
擴展區 C	1	1
擴展區 D	-	-
擴展區 E	-	-
兼容漢字區	12	12
總數	5 009	5 031

- 2.6 現有的電腦平台大都支援 ISO/IEC 10646。我們建議大五碼系統的使用者和管理者，盡快把這些系統更新至可支援 ISO/IEC 10646，以便能更快更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》

- 3.3 ISO/IEC 10646 國際編碼標準(#1)的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是中國、日本、韓國及其他國家和地區經認同規則整合後所採納的漢字字符的統稱。最早的表意文字來源主要是中國、日本和韓國，因此“中日韓”名稱沿用至今。本文也簡稱為“漢字”。

[#1 為達到推廣標準化，國際標準化組織在以下網頁免費提供一些刊載 ISO 標準的文件，當中包括 ISO/IEC 10646 國際編碼標準：<http://standards.iso.org/ittf/PubliclyAvailableStandards/>。]

- 3.4 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。
- 3.5 ISO/IEC 10646 國際編碼標準版本 ISO/IEC 10646:2014 已包含漢字逾 80 000 個，為電腦平台提供整體而全面的支援。一般而言，在實際的應用上，支持逾 80 000 個的字集，不僅會增加開發時間和費用，也未必符合個別地區的用字習慣。所以字體或輸入法開發商會因應個別地區的用字要求，選取適量字符開發支援相關產品。
- 3.6 中日韓統一表意文字區區內每一個碼位以多個字形並列的形式來支持和定義某一個國家或地區所需的漢字，香港使用的漢字羅列於 H-列。橫向擴展是指在 ISO/IEC 10646 已收納的字符上增補某一國家或地區所使用的字形和字源資料，藉此反映該國家或地區所使用的漢字的情況。
- 3.7 若有新的《香港增補字符集》字符加入 ISO/IEC 10646，該新字的字

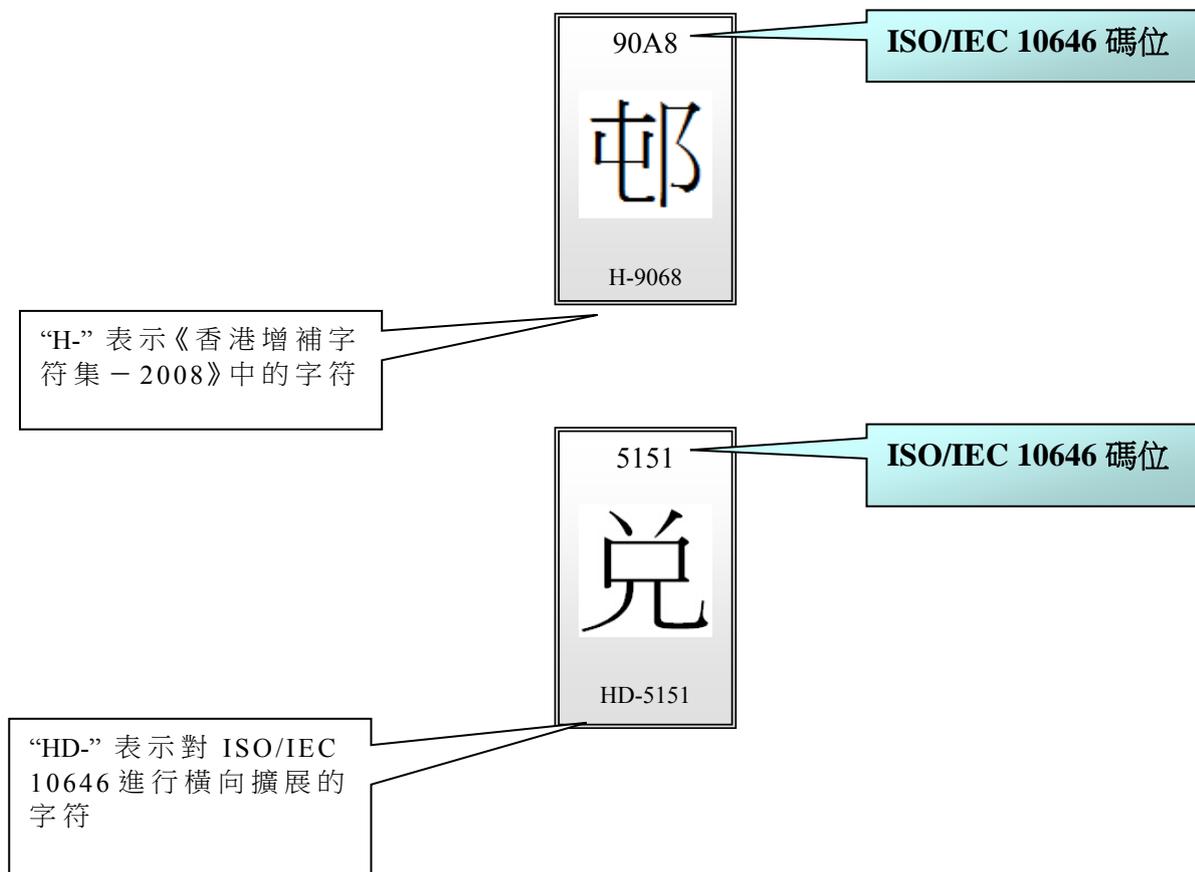
源資料會以“HC-0001” - “HC-9999”的形式以表示該新字是縱向擴展的《香港增補字符集》字符。目前 ISO/IEC 10646 並沒有以“HC-”定義的《香港增補字符集》字符。

3.8 《香港增補字符集－2015》增收了 22 個已收納在 ISO/IEC 10646 的字符，以“HD-HHHH”（“HHHH”表示是該字的 ISO/IEC 10646 碼位）表示這些字是用橫向擴展的《香港增補字符集》字符，藉此識別以縱向擴展的《香港增補字符集》字符。

3.9 在 ISO/IEC 10646 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，字源資料的形式是“HB1-XXXX”和“HB2-XXXX”，分別代表大五碼的常用字集和次常用字集，“XXXX”是對應字符在大五碼的編碼。（#2）

[#2 有關大五碼和 Unicode 的字符對應資料，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》字符表中一般方格的說明



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

附表二：橫向擴展的《香港增補字符集》字符

在《香港電腦漢字參考字形》(＃)的編訂過程中，找到 22 香港慣用的字形與大五碼所取的字形有不同之處，例如大五碼內的“兌”（Big5 0xA749；U+514C），而香港慣用的字形（“兌”）並沒有收納在《香港增補字符集－2008》內。因此中文界面諮詢委員會(中諮會)決定將這些香港慣用的漢字加入《香港增補字符集－2015》，並通過 ISO/IEC 10646 橫向擴展的形式來確認這些字符為香港慣用的。該 22 個字列在下表。

序號	大五碼字形 ISO/IEC 10646 編碼 (大五碼內碼)	香港慣用字形 橫向擴展的 ISO/IEC 10646 編碼
1	告 U+544A (0xA769)	告 U+543F
2	兌 U+514C (0xA749)	兌 U+5151
3	悅 U+6085 (0xAEAE)	悅 U+60A6
4	掙 U+6329 (0xD1BE)	掙 U+635D
5	斂 U+6553 (0xD5A7)	斂 U+655A
6	稅 U+68B2 (0xD5BF)	稅 U+68C1
7	浼 U+6D97 (0xD258)	浼 U+6D9A
8	稅 U+7A05 (0xB57C)	稅 U+7A0E
9	脫 U+812B (0xB2E6)	脫 U+8131

序號	大五碼字形 ISO/IEC 10646 編碼 (大五碼內碼)	香港慣用字形 橫向擴展的 ISO/IEC 10646 編碼
10	蛻 U+86FB (0xB8C0)	蛻 U+8715
11	說 U+8AAA (0xBBA1)	說 U+8AAC
12	銳 U+92B3 (0xBE55)	銳 U+92ED
13	閱 U+95B1 (0xBE5C)	閱 U+95B2
14	媪 U+5ABC (0xB6FE)	媪 U+5AAA
15	愠 U+614D (0xB759)	愠 U+6120
16	氫 U+6C33 (0xBA72)	氫 U+6C32
17	熅 U+7185 (0xE2BE)	熅 U+7174
18	緼 U+7E15 (0xEAD5)	緼 U+7DFC
19	膾 U+8183 (0xE3A6)	膾 U+817D
20	蘊 U+860A (0xC4AD)	蘊 U+85F4
21	輻 U+8F40 (0xEEC1)	輻 U+8F3C
22	醞 U+919E (0xC1DF)	醞 U+9196

[#有關《香港電腦漢字參考字形》的更多資料，請參閱該文件:

http://www.ogcio.gov.hk/tc/business/tech_promotion/ccli/download_area/ °]

附表三：《香港增補字符集－2015》的兼容字符

《香港增補字符集－2015》包含 12 個中日韓兼容漢字，與之認同的字符列在下表以供參考。

序號	《香港增補字符集－2015》兼容漢字	認同的字符
1	龜 F907	龜 9F9C
2	勇 2F825	勇 52C7
3	𠵼 2F83B	𠵼 5406
4	𠵽 2F840	𠵽 54A2
5	𠵾 2F878	𠵾 5C6E
6	𠵿 2F894	𠵿 5F22
7	茲 2F8A6	茲 6148
8	晉 2F8CD	晉 6649
9	𠵿 2F994	芳 82B3
10	夔 2F9B2	夔 456B
11	𧈧 2F9BC	𧈧 8728
12	貫 2F9D4	貫 8CAB