

無線流動裝置對《香港增補字符集》的支援

香港的無線流動裝置(包括手提電話、個人數碼助理等)用戶和普羅大眾，大部分都認為無線流動裝置支援《香港增補字符集》的字符，是理所當然的事。他們無須細想，便在手提電話輸入這些廣東口語：

“今晚係咪喺深水埗黃金嗰間譚仔等？”

“噉食完飯做咩好？”

透過短訊和即時通訊，這些生動的口語，經常會顯示在無線流動裝置和桌上電腦的屏幕。本文嘗試就無線流動裝置支援《香港增補字符集》的現況，進行有系統的研究。

《香港增補字符集》

《香港增補字符集》原為《台灣大五碼字符集》的補充字符集。在統一碼和通用字元(通用字元為 ISO 10646 國際編碼標準的字符集)獲廣泛認同和採納前，分別代表繁體和簡化中文字(或漢字)的大五碼(13 053 個漢字)和國家標準碼 GB2312 (6 763 個漢字)這兩個字符集，在電腦世界非常流通。國家標準碼 GB2312 是中國制訂的國家標準，而大五碼則是由台灣五家電子／電腦業公司制訂的標準。

基於獨特的歷史背景，香港和許多海外華人社會沿用廣東話(一種華南方言)作為日常口語，並以繁體字書寫(相對於簡化字)，因而對中文字符集的要求有別於內地和台灣。儘管大五碼包含大部分常用的繁體字，並因此用作電腦預設的中文字符集，但仍欠缺許多本地的常用字。

作為大五碼的補充

大五碼系統容許增添用戶界定的文字。在大五碼系統，由十六進位的 8140 至 FEFE 為無編配字符的空間，作為使用者造字區。

碼位	字符	指定用途
8140 - 8DFE	2 041	使用者造字區 3
8E40 - A0FE	2 983	使用者造字區 2

A140 - A3FE	471	大五碼符號
A440 - C67E	5 401	大五碼第 1 級漢字
C6A1 - C8FE	408	廠商造字區：為倚天廠商字集所擁有
C940 - F9D5	7 652	大五碼第 2 級漢字
F9D6 - F9FE	41	廠商造字區：為倚天廠商字集所擁有
FA40 - FEFE	785	使用者造字區 1

香港多家公司開發這些使用者造字區，製作了多套廠商專用，但互不兼容的增補字符集：華康廠商字集 A(784 個漢字)、華康廠商字集 B(665 個漢字加 746 個符號)，以及蒙納廠商字集(471 個漢字)。

由於政府內部和市民對電子通訊以及電腦版中文刊物的需求日增，為協調各種不同的工作，香港政府於 1995 年開發《政府通用字庫》，其後於 1999 年修訂並命名為《香港增補字符集》。《香港增補字符集》廣受業界人士和市民採用，並獲國際標準化組織(International Organization for Standardization)和出版界認同。

關於統一碼

《政府通用字庫》起初屬大五碼碼頁的擴展部分，加入 3 049 個本港常用、但不存在於大五碼的古字和新字。不過，其模式後來由本地化轉為國際化。由於統一碼愈來愈普及，大家逐漸以通用字元取代大五碼，作為參考的基準。《香港增補字符集》內的字符適用於大五碼和 ISO 10646 國際編碼標準。在最新版本的《香港增補字符集—2004》中，其 ISO 10646 國際編碼標準涵蓋的範圍，較以往的版本更廣泛和顯著。

自 1991 年開始，開發統一碼的 Unicode Consortium(Unicode 學術學會)和國際標準化組織已合力統一工作，令統一碼標準和 ISO 10646 國際編碼標準的主要版本有對應編碼。統一碼標準編有大量字符，包含全球所需的字符。通用字元的編碼空間遠超大五碼，但與大五碼同樣設有私人使用區(和私人使用平面)，以供個人使用。預設區(或字區)也包含餘下的碼位，作日後擴展之用。以下為通用字元的部分字區和平面實例：

碼位	字符	字區
0000 - FFFD	65 536	基本多文種平面 (或平面 0)
0000 - 1FFF		通用文字
2000 - 2DFF		通用符號
2E00 - 2FFF		康熙字典部首
3000 - 33FF		中日韓符號

3400 - 4DBF		中日韓等同表意文字擴展區 A
4E00 - 9FFF		中日韓等同表意文字
E000 - F8FF	6 400	私人使用區
F900 - FAFF	512	中日韓兼容表意文字(基本多文種平面)
10000 - 1FFFD	65 536	增補多文種平面(或平面 1)
20000 - 2FFFD	65 536	增補表意文字平面(或平面 2)
20000 - 2A6DF	42 720	中日韓等同表意文字擴展區 B
2F800 - 2FA1D	542	中日韓兼容表意文字
E0000 - EFFFFD	65 536	增補特定用途平面 (或平面 14)
F0000 - 10FFFFD	131 068	私人使用平面(平面 15 及 16)

香港特別行政區政府一直與“中文界面諮詢委員會”合作，通過檢討和增收字符，改善和擴展《香港增補字符集》，並向國際標準化組織轄下的表意文字小組提交新增字符，以便最終納入 ISO 10646 國際編碼標準各版本內。近年，《香港增補字符集》已多次修訂：由 3 049 個字符的《政府通用字庫》修訂為《香港增補字符集—1999》(4 702 個字符)、《香港增補字符集—2001》(加入 116 個字符，合共 4 818 個字符)，以及在 2005 年推出的《香港增補字符集—2004》(再加入 123 個字符，合共 4 941 個字符)。

《香港增補字符集》版本	字符數目
《政府通用字庫》(1995)	3 049
《香港增補字符集—1999》	4 702
《香港增補字符集—2001》	4 818
《香港增補字符集—2004》	4 941

在《香港增補字符集》不斷增長之餘，因應統一碼和 ISO 10646 國際編碼標準的多次修訂，通用字元也有所增加。尚未納入 ISO 10646 國際編碼標準的《香港增補字符集》字符通過上述提交程序，可加入 ISO 10646 國際編碼標準。在《政府通用字庫》以前推出的 ISO/IEC 10646-1:1993(與統一碼版本 1.1 對應)，已包含了約 2 000 個《香港增補字符集》字符，以下為一些例子：

《香港增補字符集》字符	碼位
㗎	U+5497
咩	U+54A9
咪	U+54AA
㗎	U+55BA
㗎	U+55F0
嘅	U+5606
嘞	U+561E

噉	U+5649
嚟	U+569F
埗	U+57D7
邨	U+90A8
鯪	U+9C02

餘下的字符(例如“𦉳”)則編入私人使用區(基本多文種平面)內，一經接納，會遷至相關的字區。《香港增補字符集》的字符，其後陸續收納在ISO 10646國際編碼標準各版本內。

ISO 10646 國際編碼標準	統一碼	《香港增補字符集》字符
ISO/IEC 10646-1:1993	版本 1.1	2 000+
ISO/IEC 10646-1:2000	版本 3.0	3 054=4 702 - 1 648 (《香港增補字符集－1999》)
ISO/IEC 10646:2000 附加 ISO/IEC 10646-2:2001	版本 3.1	4 783 =4 818 - 35(《香港增補字符集－2001》)
ISO/IEC 10646:2003 附加第一修訂版	版本 4.1	4 941 (《香港增補字符集－2004》)
ISO/IEC 10646:2003 附加第一、第二修訂版及部分第三修訂版	版本 5.0	—

自 ISO/IEC 10646-1:2000(統一碼版本 3.0)、ISO/IEC 10646-2:2001(統一碼版本 3.1)和 ISO/IEC 10646:2003 附加第一修訂版(統一碼版本 4.1)初次推出¹後，已分別涵蓋 3 054 個、4 783 個和 4 941 個《香港增補字符集》字符。

流動裝置的字符集

流動裝置屬新興的電腦設備，其發展歷史甚短，因而從開始已大致能配合統一碼。例如短訊服務(環球流動通訊系統規格中很受歡迎的文字訊息平台)就適用於通用字元-2 內的所有非美國信息交換標準碼字符(屬統一碼 2 字節編碼形式)。如要知道某種裝置對《香港增補字符集》的支援情況，只需了解該裝置支援哪個統一碼版本就可以了。

流動裝置原則上支援統一碼，但並非支援所有統一碼的字符。基於實際需要，統一碼字符又有以下各子集：

¹ 在 ISO/IEC 10646 國際編碼標準各版本初次推出後，所收錄的《香港增補字符集》字符數目持續增長，主要原因是這些版本的字符陸續納入《香港增補字符集》的新版本內。有關情況，可參考 <http://mobile.hkwdc.org/hkwdc/utest.html> 所載列的圖表。

《統一碼子集》
《最小歐洲子集》
《多文種歐洲子集》(包括：《多文種歐洲子集-1》、《多文種歐洲子集-2》和《多文種歐洲子集-3》)
《視窗字形列表 4》
《國際表意文字子集》

有關方面已設立網頁，通過裝置內的瀏覽器，測試裝置本身可以使用的字符 (<http://mobile.hkwdc.org/hkwdc/utest.html>)。結果顯示，2007 年以前推出的大部分手機，只支援《香港增補字符集》某些字符。這些字符多數屬於《香港增補字符集－1999》，另有少量為《香港增補字符集－2001》所增收的 116 個字符，而《香港增補字符集－2004》所增收的 123 個字符更是絕無僅有。

下面以一些受歡迎的流動裝置為例，比較它們在顯示‘儻’（‘俊’的異體字）和‘𠵼’這兩個《香港增補字符集－1999》字符的情況。在納入 ISO 10646 國際編碼標準(ISO/IEC 10646-1:2000 和 ISO/IEC 10646-2:2001)以前，這兩個字符都載於私人使用區，碼位分別為 U+E028 和 U+F442，其後獲編配為 U+347A（中日韓等同表意文字擴展區 A）和 U+2512B（中日韓等同表意文字擴展區 B）。

流動裝置 (推出日期)	儻 (U+E028)	儻 (U+347A)	𠵼 (U+F442)	𠵼 (U+2512B)
諾基亞 6108(2003 年 5 月)	□	儻	𠵼	𥄫
諾基亞 6230(2003 年 10 月)	□	□	□	𥄫
諾基亞 7610(2004 年 3 月)	[空白]	儻	𠵼	𥄫
諾基亞 6630(2004 年 6 月)	□	儻	𠵼	𥄫
諾基亞 N73(2006 年 4 月)	儻	儻	𠵼	𥄫
諾基亞 N95(2006 年 9 月)	儻	儻	𠵼	[空白]

以上例子可以反映一般流動裝置的情況。除中日韓等同表意文字的主編碼頁外，擴展區 A 是最早獲支援的。鑑於流動裝置日漸複雜，反向兼容應運而生，以便準確顯示私人使用區的碼位。在 2006 年年中左右，當超過 2 字節的編碼獲承認為字符後(而非只屬數值，不能轉為可顯示的字符)，擴展區 B 的字符才獲支援。如無合適的字型支援，這些字符則視作空白(或如上表所示的空格)。

下表(摘自 <http://mobile.hkwdc.org/hkwdc/utest.html>)顯示，《香港增補字符集》的字符，已分期納入 ISO 10646 國際編碼標準內。
(下表的第一、二行與英文版不符)

ISO 10646 國際	《香港增補字符集》字符數目
--------------	---------------

編碼標準的字區	ISO/IEC 10646-1:1993	ISO/IEC 10646-1:2000	ISO/IEC 10646-2:2001	ISO/IEC 10646:2003 及第一修訂版
基本多文種平面				
字母區 (0000-1FFF)	123	123	123	127
通用符號區 (2000-2DFF)	69	69	69	71
康熙字典部首區 (2E00-2FFF)	不適用	29	29	29
中日韓符號區 (3000-33FF)	180	180	180	196
中日韓表意 文字擴展區 A (3400-4DBF)	不適用	<ul style="list-style-type: none"> • 562 (《香港增補字符集-2004》) • 511 (《香港增補字符集-2001》) • 498 (《香港增補字符集-1999》) 	511	562
中日韓表意 文字區 (4E00-9FFF)	<ul style="list-style-type: none"> • 2 241 (《香港增補字符集-2004》) • 2 212 (《香港增補字符集-2001》) • 2 148 (《香港增補字符集-1999》) 	<ul style="list-style-type: none"> • 2 241 (《香港增補字符集-2004》) • 2 212 (《香港增補字符集-2001》) • 2 148 (《香港增補字符集-1999》) 	2 212	2 255
... ..				
私人使用區 (E000-F8FF)	<ul style="list-style-type: none"> • 2 320 (《香港增補字符集-2004》) • 2 226 (《香港增補字符集-2001》) • 2 175 (《香港增補字符集-1999》) 	<ul style="list-style-type: none"> • 1 729 (《香港增補字符集-2004》) • 1 686 (《香港增補字符集-2001》) • 1 648 (《香港增補字符集-1999》) 	35	0
兼容字符區 (基本多文種 平面) (F900-FAFF)	<ul style="list-style-type: none"> • 8 (《香港增補字符集-2001/4》) • 7 (《香港增補字符集-1999》) 	<ul style="list-style-type: none"> • 8 (《香港增補字符集-2001/4》) • 7 (《香港增補字符集-1999》) 	8	8
增補表意文字平面				
中日韓表意 文字擴展區 B (20000-2A6DF)	不適用		1 640	1 682
... ..				

兼容字符區 (增補表意文字 平面) (2F800-2FA1D)	不適用	11	11
------------------------------------------	-----	----	----

自從 1993 年 ISO 10646 國際編碼標準首個版本推出以來，已有超過 2 000 個《香港增補字符集》字符納入其中。

於 2000 年和 2001 年推出的新版本標準，分別開拓了兩個擴展區：基本多文種平面內的擴展區 A，以及增補表意文字平面內的擴展區 B，足以為 6 592 個和 42 720 個新增收的漢字(中日韓等同表意文字)提供碼位。

首先，未曾納入 ISO 10646 國際編碼標準的《香港增補字符集》字符，獲編入中日韓表意文字擴展區 A 內。該區用以匯集不常用的中日韓表意文字字符(例如字符‘僂’)。《香港增補字符集》大部分的新增收字符，已於 2001 年納入中日韓表意文字擴展區 B，包括常見的廣東話字符，例如‘𦉳’(由 U+F442 遷至 U+2512B)和‘𦉴’(由 U+EB78 遷至 U+20779)。

由於短訊採用通用字元-2 為非美國信息交換標準碼字符編碼，所以中日韓表意文字擴展區 B 內的《香港增補字符集》字符便不能使用，這是通用字元-2 只能顯示平面 0(基本多文種平面)字符的固有限制。

本港出售的手機和流動裝置，普遍支援《香港增補字符集－1999》的大部分字符。最新的裝置雖可支援擴展區 B 的字符(於《香港增補字符集－2001》和《香港增補字符集－2004》新加入的字符)，但或會欠缺顯示字符所需的字型。然而，即使是桌上電腦，對《香港增補字符集－2001》或《香港增補字符集－2004》作出支援，仍算是初次涉獵中日韓表意文字擴展區 B。該區之前從未為鑄字廠或造字商所開拓。隨着記憶組件愈見便宜和處理能力不斷上升，假以時日流動裝置定會達到桌上電腦現時支援《香港增補字符集》的水平。

