

无线流动装置对《香港增补字符集》的支援

香港的无线流动装置(包括手提电话、个人数码助理等)用户和普罗大众, 大部分都认为无线流动装置支持《香港增补字符集》的字符, 是理所当然的事。他们无须细想, 便在手提电话输入这些广东口语:

“今晚系咪喺深水埗黄金嗰间谭仔等?”

“噉食完饭做咩好?”

透过短讯和实时通讯, 这些生动的口语, 经常会显示在无线流动装置和桌上计算机的屏幕。本文尝试就无线流动装置支持《香港增补字符集》的现况, 进行有系统的研究。

《香港增补字符集》

《香港增补字符集》原为《台湾大五码字符集》的补充字符集。在统一码和通用字符(通用字符为 ISO 10646 国际编码标准的字符集)获广泛认同和采纳前, 分别代表繁体和简化中文字(或汉字)的大五码(13 053 个汉字)和国家标准码 GB2312 (6 763 个汉字)这两个字符集, 在计算机世界非常流通。国家标准码 GB2312 是中国制订的国家标准, 而大五码则是由台湾五家电子/计算机业公司制订的标准。

基于独特的历史背景, 香港和许多海外华人社会沿用广东话(一种华南方言)作为日常口语, 并以繁体字书写(相对于简化字), 因而对中文字符集的要求有别于内地和台湾。尽管大五码包含大部分常用的繁体字, 并因此用作计算机预设的中文字符集, 但仍欠缺许多本地的常用字。

作为大五码的补充

大五码系统容许增添用户界定的文字。在大五码系统, 由十六进制的 8140 至 FEFE 为无编配字符的空间, 作为使用者造字区。

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

A140 - A3FE	471	大五码符号
A440 - C67E	5 401	大五码第 1 级汉字
C6A1 - C8FE	408	厂商造字区：为倚天厂商字集所拥有
C940 - F9D5	7 652	大五码第 2 级汉字
F9D6 - F9FE	41	厂商造字区：为倚天厂商字集所拥有
FA40 - FEFE	785	使用者造字区 1

香港多家公司开发这些使用者造字区，制作了多套厂商专用，但互不兼容的增补字符集：华康厂商字集 A(784 个汉字)、华康厂商字集 B(665 个汉字加 746 个符号)，以及蒙纳厂商字集(471 个汉字)。

由于政府内部和市民对电子通讯以及计算机版中文刊物的需求日增，为协调各种不同的工作，香港政府于 1995 年开发《政府通用字库》，其后于 1999 年修订并命名为《香港增补字符集》。《香港增补字符集》广受业界人士和市民采用，并获国际标准化组织(International Organization for Standardization)和出版界认同。

关于统一码

《政府通用字库》起初属大五码码页的扩展部分，加入 3 049 个本港常用、但不存在于大五码的古字和新字。不过，其模式后来由本地化转为国际化。由于统一码愈来愈普及，大家逐渐以通用字符取代大五码，作为参考的基准。《香港增补字符集》内的字符适用于大五码和 ISO 10646 国际编码标准。在最新版本的《香港增补字符集—2004》中，其 ISO 10646 国际编码标准涵盖的范围，较以往的版本更广泛和显著。

自 1991 年开始，开发统一码的 Unicode Consortium(Unicode 学术学会)和国际标准化组织已合力统一工作，令统一码标准和 ISO 10646 国际编码标准的主要版本有对应编码。统一码标准编有大量字符，包含全球所需的字符。通用字符的编码空间远超大五码，但与大五码同样设有私人使用区(和私人使用平面)，以供个人使用。预设区(或字区)也包含余下的码位，作日后扩展之用。以下为通用字符的部分字区和平面实例：

码位	字符	字区
0000 - FFFD	65 536	基本多文种平面（或平面 0）
0000 - 1FFF		通用文字
2000 - 2DFF		通用符号
2E00 - 2FFF		康熙字典部首

3000 - 33FF		中日韩符号
3400 - 4DBF		中日韩等同表意文字扩展区 A
4E00 - 9FFF		中日韩等同表意文字
E000 - F8FF	6 400	私人使用区
F900 - FAFF	512	中日韩兼容表意文字(基本多文种平面)
10000 - 1FFFD	65 536	增补多文种平面(或平面 1)
20000 - 2FFFD	65 536	增补表意文字平面(或平面 2)
20000 - 2A6DF	42 720	中日韩等同表意文字扩展区 B
2F800 - 2FA1D	542	中日韩兼容表意文字
E0000 - EFFFFD	65 536	增补特定用途平面 (或平面 14)
F0000 - 10FFFFD	131 068	私人使用平面(平面 15 及 16)

香港特别行政区政府一直与“中文界面咨询委员会”合作，通过检讨和增收字符，改善和扩展《香港增补字符集》，并向国际标准化组织辖下的表意文字小组提交新增字符，以便最终纳入 ISO 10646 国际编码标准各版本内。近年，《香港增补字符集》已多次修订：由 3 049 个字符的《政府通用字库》修订为《香港增补字符集—1999》(4 702 个字符)、《香港增补字符集—2001》(加入 116 个字符，合共 4 818 个字符)，以及在 2005 年推出的《香港增补字符集—2004》(再加入 123 个字符，合共 4 941 个字符)。

《香港增补字符集》版本	字符数目
《政府通用字库》(1995)	3 049
《香港增补字符集—1999》	4 702
《香港增补字符集—2001》	4 818
《香港增补字符集—2004》	4 941

在《香港增补字符集》不断增长之余，因应统一码和 ISO 10646 国际编码标准的多次修订，通用字符也有所增加。尚未纳入 ISO 10646 国际编码标准的《香港增补字符集》字符通过上述提交程序，可加入 ISO 10646 国际编码标准。在《政府通用字库》以前推出的 ISO/IEC 10646-1:1993(与统一码版本 1.1 对应)，已包含了约 2 000 个《香港增补字符集》字符，以下为一些例子：

《香港增补字符集》字符	码位
㗎	U+5497
咩	U+54A9
咪	U+54AA
喺	U+55BA
嗰	U+55F0

嘅	U+5606
嘞	U+561E
噉	U+5649
嚟	U+569F
埗	U+57D7
邨	U+90A8
鰌	U+9C02

余下的字符(例如“𨳐”)则编入私人使用区(基本多文种平面)内，一经接纳，会迁至相关的字区。《香港增补字符集》的字符，其后陆续收纳在ISO 10646国际编码标准各版本内。

ISO 10646 国际编码标准	统一码	《香港增补字符集》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 (《香港增补字符集—1999》)
ISO/IEC 10646:2000 附加 ISO/IEC 10646-2:2001	版本 3.1	4 783 =4 818 - 35(《香港增补字符集—2001》)
ISO/IEC 10646:2003 附加第一修订版	版本 4.1	4 941 (《香港增补字符集—2004》)
ISO/IEC 10646:2003 附加第一、第二修订版及部分第三修订版	版本 5.0	—

自 ISO/IEC 10646-1:2000(统一码版本 3.0)、ISO/IEC 10646-2:2001(统一码版本 3.1)和 ISO/IEC 10646:2003 附加第一修订版(统一码版本 4.1)初次推出¹后，已分别涵盖 3 054 个、4 783 个和 4 941 个《香港增补字符集》字符。

流动装置的字符集

流动装置属新兴的计算机设备，其发展历史甚短，因而从开始已大致能配合统一码。例如短讯服务(环球流动通讯系统规格中很受欢迎的文字讯息平台)就适用于通用字符-2 内的所有非美国信息交换标准码字符(属统一码 2 字节编码形式)。如要知道某种装置对《香港增补字符集》的支持情况，只需了解该装置支持哪个统一码版本就可以了。

¹ 在 ISO/IEC 10646 国际编码标准各版本初次推出后，所收录的《香港增补字符集》字符数目持续增长，主要原因是这些版本的字符陆续纳入《香港增补字符集》的新版本内。有关情况，可参考 <http://mobile.hkwdc.org/hkwdc/utest.html> 所载列的图表。

流动装置原则上支持统一码，但并非支持所有统一码的字符。基于实际需要，统一码字符又有以下各子集：

《统一码子集》
《最小欧洲子集》
《多文种欧洲子集》(包括：《多文种欧洲子集-1》、《多文种欧洲子集-2》和《多文种欧洲子集-3》)
《窗口字形列表 4》
《国际表意文字子集》

有关方面已设立网页，通过装置内的浏览器，测试装置本身可以使用的字符 (<http://mobile.hkwdc.org/hkwdc/utest.html>)。结果显示，2007 年以前推出的大部分手机，只支持《香港增补字符集》某些字符。这些字符多数属于《香港增补字符集—1999》，另有小量为《香港增补字符集—2001》所增收的 116 个字符，而《香港增补字符集—2004》所增收的 123 个字符更是绝无仅有。

下面以一些受欢迎的流动装置为例，比较它们在显示‘儻’(‘俊’的异体字)和‘𠵼’这两个《香港增补字符集—1999》字符的情况。在纳入 ISO 10646 国际编码标准 (ISO/IEC 10646-1:2000 和 ISO/IEC 10646-2:2001) 以前，这两个字符都载于私人使用区，码位分别为 U+E028 和 U+F442，其后获编配为 U+347A (中日韩等同表意文字扩展区 A) 和 U+2512B (中日韩等同表意文字扩展区 B)。

流动装置 (推出日期)	儻 (U+E028)	儻 (U+347A)	𠵼 (U+F442)	𠵼 (U+2512B)
诺基亚 6108(2003 年 5 月)	□	儻	𠵼	𥄫
诺基亚 6230(2003 年 10 月)	□	□	□	𥄫
诺基亚 7610(2004 年 3 月)	[空白]	儻	𠵼	𥄫
诺基亚 6630(2004 年 6 月)	□	儻	𠵼	𥄫
诺基亚 N73(2006 年 4 月)	儻	儻	𠵼	𥄫
诺基亚 N95(2006 年 9 月)	儻	儻	𠵼	[空白]

以上例子可以反映一般流动装置的情况。除中日韩等同表意文字的主编码页外，扩展区 A 是最早获支持的。鉴于流动装置日渐复杂，反向兼容应运而生，以便准确显示私人使用区的码位。在 2006 年年中左右，当超过 2 字节的编码获承认为字符后(而非只属数值，不能转为可显示的字符)，扩展区 B 的字符才获支持。如无合适的字型支持，这些字符则视作空白(或如上表所示的空格)。

下表(摘自 <http://mobile.hkwdc.org/hkwdc/utest.html>)显示，《香港增补字符

集》的字符，已分期纳入 ISO 10646 国际编码标准内。

(下表的第二、三行与英文版不符)

ISO 10646 国际 编码标准的 字区	《香港增补字符集》字符数目			
	ISO/IEC 10646-1:1993	ISO/IEC 10646-1:2000	ISO/IEC 10646-2:2001	ISO/IEC 10646:2003 及第一修订版
基本多文种平面				
字母区 (0000-1FFF)	123	123	123	127
通用符号区 (2000-2DFF)	69	69	69	71
康熙字典部首区 (2E00-2FFF)	不适用	29	29	29
中日韩符号区 (3000-33FF)	180	180	180	196
中日韩表意 文字扩展区 A (3400-4DBF)	不适用	<ul style="list-style-type: none"> • 562 (《香港增补字符集-2004》) • 511 (《香港增补字符集-2001》) • 498 (《香港增补字符集-1999》) 	511	562
中日韩表意 文字区 (4E00-9FFF)	<ul style="list-style-type: none"> • 2 241 (《香港增补字符集-2004》) • 2 212 (《香港增补字符集-2001》) • 2 148 (《香港增补字符集-1999》) 	<ul style="list-style-type: none"> • 2 241 (《香港增补字符集-2004》) • 2 212 (《香港增补字符集-2001》) • 2 148 (《香港增补字符集-1999》) 	2 212	2 255
... ..				
私人使用区 (E000-F8FF)	<ul style="list-style-type: none"> • 2 320 (《香港增补字符集-2004》) • 2 226 (《香港增补字符集-2001》) • 2 175 (《香港增补字符集-1999》) 	<ul style="list-style-type: none"> • 1 729 (《香港增补字符集-2004》) • 1 686 (《香港增补字符集-2001》) • 1 648 (《香港增补字符集-1999》) 	35	0
兼容字符区 (基本多文种 平面) (F900-FAFF)	<ul style="list-style-type: none"> • 8 (《香港增补字符集-2001/4》) • 7 (《香港增补字符集-1999》) 	<ul style="list-style-type: none"> • 8 (《香港增补字符集-2001/4》) • 7 (《香港增补字符集-1999》) 	8	8
增补表意文字平面				

中日韩表意文字扩展区 B (20000-2A6DF)	不适用	1 640	1 682
... ..			
兼容字符区 (增补表意文字平面) (2F800-2FA1D)	不适用	11	11

自从 1993 年 ISO 10646 国际编码标准首个版本推出以来，已有超过 2 000 个《香港增补字符集》字符纳入其中。

于 2000 年和 2001 年推出的新版本标准，分别开拓了两个扩展区：基本多文种平面内的扩展区 A，以及增补表意文字平面内的扩展区 B，足以为 6 592 个和 42 720 个新增收的汉字(中日韩等同表意文字)提供码位。

首先，未曾纳入 ISO 10646 国际编码标准的《香港增补字符集》字符，获编入中日韩表意文字扩展区 A 内。该区用以汇集不常用的中日韩表意文字字符(例如字符‘僂’)。《香港增补字符集》大部分的新增收字符，已于 2001 年纳入中日韩表意文字扩展区 B，包括常见的广东话字符，例如‘𦉳’(由 U+F442 迁至 U+2512B)和‘𦉳’(由 U+EB78 迁至 U+20779)。

由于短讯采用通用字符-2 为非美国信息交换标准码字符编码，所以中日韩表意文字扩展区 B 内的《香港增补字符集》字符便不能使用，这是通用字符-2 只能显示平面 0(基本多文种平面)字符的固有限制。

本港出售的手机和流动装置，普遍支持《香港增补字符集—1999》的大部分字符。最新的装置虽可支持扩展区 B 的字符(于《香港增补字符集—2001》和《香港增补字符集—2004》新加入的字符)，但或会欠缺显示字符所需的字型。然而，即使是桌上计算机，对《香港增补字符集—2001》或《香港增补字符集—2004》作出支援，仍算是初次涉猎中日韩表意文字扩展区 B。该区之前从未为铸字厂或造字商所开拓。随着记忆组件愈见便宜和处理能力不断上升，假以时日流动装置定会达到桌上计算机现时支持《香港增补字符集》的水平。