

## 第一部分 概論

1.1 本文件旨在說明《香港增補字符集－2004》的整體編碼架構和編碼方案，並提供該字符集在 ISO/IEC 10646:2003 及第一修訂版、ISO/IEC 10646-1:2000、ISO/IEC 10646-1:1993 和大五碼的對應碼位。未包含在 ISO/IEC 10646-1:2000 和 ISO/IEC 10646-1:1993 內的《香港增補字符集－2004》字符，會編入相應 ISO 10646 國際編碼標準版本內的私人使用區內。在大五碼系統內，《香港增補字符集－2004》的字符則編入使用者造字區及廠商造字區內。《香港增補字符集－2004》與《香港增補字符集》以前的兩個版本及《政府通用字庫》完全兼容。

1.2 《香港增補字符集－2004》是一套編碼字符集，而非字形標準。

(註:有關適用於香港的電腦漢字字形原則的資料，可參考《香港電腦漢字楷體字形參考指引》及《香港電腦漢字宋體(印刷體)字形參考指引》(<http://www.info.gov.hk/digital21/unicode/structure/glyph.html>)。)

1.3 以下是一些在本文件中出現的名詞及其定義：

**反向兼容**  
(backward compatibility) 指新版標準支持舊版標準中原有的編碼。有時僅指新標準支持原標準內的所有字符。

**字區**  
(block) 相互連接並具有共同特徵的一組字符。

**字符**  
(character) 用於組織、控制或表示數據的諸元素集合中的一員。

**字符集**  
(character set) 一組已定義的字符。

碼位 (code point)	代表一個字符的二進制碼。
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字碼的對應關係。
字形 (glyph)	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。
政府通用字庫 (Government Common Character Set, GCCS)	香港政府於一九九五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。
ISO 10646	國際標準化組織制定的一套編碼字符集。建立這套標準的目的，是提供一個包含所有主要語言的單一字符集。
ISO/IEC 10646-1:1993	指於一九九三年發表的 ISO 10646 國際編碼標準，共有 20,902 個表意文字。
ISO/IEC 10646-1:2000	指於二零零零年發表的 ISO 10646 國際編碼標準。該版本是 ISO/IEC 10646-1:1993 版的擴展，新增了 6,582 個表意文字（放在擴展區 A 內）、214 個康熙部首及其他符號。
ISO/IEC 10646-2:2001	指於二零零一年發表的 ISO 10646 國際編碼標準。該版本新增了 42,711 個表意文字（放在擴展區 B 內）。

ISO/IEC 10646:2003	指於二零零四年發表的 ISO 10646 國際編碼標準。該版本合併了 ISO/IEC 10646-1:2000 和它的補充版本 ISO/IEC 10646-2:2001 而成為單一的發行本。
ISO/IEC 10646:2003 第一修訂版 (ISO/IEC 10646:2003 Amendment 1)	指於二零零四年通過的國際編碼標準 ISO/IEC 10646:2003 第一修訂版。
八位字節 (octet)	視為一體的八位字節序列。
私人使用區 (Private Use Area, PUA)	在 ISO 10646 國際編碼標準中預留給用戶及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及廠商作私人用途，不用作資料交換。
使用者造字區 (User-Defined Area, UDA)	在大五碼中預留給用戶自行定義字符的字區。原則上，大五碼的使用者造字區只供個別用戶作私人用途，不用作資料交換。
等同 (unification)	把一個碼位編配給兩個或以上看來不同但實際上是異體字符關係的圖形符號。有關符號代表同一數據元素，因此只選擇其中一個作為代表。
廠商造字區 (Vendor-Defined Area, VDA)	在大五碼中預留給廠商自行定義字符的字區。有些系統容許用戶使用這些字區作私人用途。