

香港電腦漢字宋體(印刷體)字形參考指引

Reference Guide on Song Style (Print Style) Character Glyphs for Chinese Computer Systems in Hong Kong

1. 主題內容

- 1.1 本指引的目的是提供一套適合香港使用的宋體(印刷體)電腦漢字字形原則供業界參考。
- 1.2 本指引適用於 ISO/IEC 10646 國際標準中的基本多文種平面(BMP)、A 擴充集(Extension A) 及《香港增補字符集 — 2001》(Hong Kong Supplementary Character Set - 2001) 內的漢字, 但不適用於 ISO/IEC 10646 國際標準所包括的內地規範簡化字及僅在日本、韓國、越南或新加坡地區使用的漢字¹。
- 1.3 本指引以《常用字字形表(二零零零年修訂本)》(見下文 2.2.2 段)的字形為基礎, 把組成本指引適用範圍內所有漢字的部件一一列出(部件拆分原則見下文第 4 段), 並列舉合適的漢字作為例子, 藉此闡明適合香港使用的字形原則。
- 1.4 本指引只描述漢字部件的筆畫結構, 並不對漢字或其部件之間的大小和比例構成限制。

2. 參考標準與文獻

2.1 標準

- 2.1.1 ISO/IEC 10646-1:2000(E), *International Standard: Information Technology-Universal Multiple-Octet Coded Character Set (UCS) - Part 1: Architecture and Basic Multilingual Plane*, ISO/IEC, 2000.
- 2.1.2 GB/T 12200.2-94 《漢語信息處理詞彙 02 部分: 漢語和漢字》, 國家技術監督局, 中國標準出版社, 1994 年 12 月。
- 2.1.3 GB 13000.1-93 《信息技術 通用多八位編碼字符集(UCS), 第一部分 體系結構與基本多文種平面》, 國家技術監督局, 中國標準出版社, 1993 年 12 月。
- 2.1.4 GB/T 14717-93 《信息交換用漢字 128X128 點陣宋體字模集及數據集》, 國家技術監督局, 中國標準出版社, 1993 年。
- 2.1.5 GB/T 14719-93 《信息交換用漢字 256X256 點陣宋體字模集及數據集》, 國家技術監督局, 中國標準出版社, 1993 年。
- 2.1.6 GB 12200.1-90 《漢語信息處理詞彙 01 部分: 基本術語》, 國家技術監督局, 中國標準出版社, 1990 年 2 月。

¹ 本指引建議業界參照 ISO/IEC 10646 標準中 G 列字形製造此類漢字。

2.2 文獻

- 2.2.1 《香港增補字符集 — 2001》，香港特別行政區政府，2001年12月。
- 2.2.2 《常用字字形表（二零零零年修訂本）》，李學銘主編，香港教育學院，2000年。
- 2.2.3 *The Unicode Standard Version 3.0*, The Unicode Consortium, The Unicode Inc., 2000.
- 2.2.4 〈漢字筆畫定量研究〉，馬顯彬，《語文建設通訊(香港)》，第63期，2000年4月。
- 2.2.5 《香港增補字符集》，香港特別行政區政府，1999年9月。
- 2.2.6 《國字標準字體宋體母稿 — 教育部字序》，台灣教育部，1998年2月。
- 2.2.7 GF3001-1997《國家語言文字工作委員會 語言文字規範 信息處理用 GB13000.1 字符集 漢字部件規範》，國家語言文字工作委員會，1997年12月。
- 2.2.8 《康熙字典》，中華書局影印本，1997年2月。
- 2.2.9 《現代漢字學綱要》，蘇培成，北京大學出版社，1994年10月。
- 2.2.10 《現代漢語通用字 — 筆順規範》，國家語言文字工作委員會標準化工作委員會，語文出版社，1997年8月。
- 2.2.11 《簡化字總表》，中國文字改革委員會，語文出版社，1986年。
- 2.2.12 《漢語大字典》，漢語大字典編輯委員會，湖北辭書出版社、四川辭書出版社，1986年10月。
- 2.2.13 《電腦用中文字型與字碼對照表》²（技術通報 C-26），財團法人資訊工業策進會，1984年5月。
- 2.2.14 《常用國字標準字體表》，台灣教育部，1982年9月。
- 2.2.15 《香港電腦漢字楷體字形參考指引》，香港特別行政區政府，日期待定。

3. 術語

- 3.1 統一的中日韓文字 CJK Unified Ideographs (ISO/IEC 10646 《信息技術 通用多八位編碼字符集 (UCS)》)
ISO/IEC 10646 國際標準包括的中國、日本、韓國及其他國家和地區經認同規則整合後所採納的字符的統稱。
- 3.2 漢字 Chinese Character / Hanzi (GB 12200.1-90 《漢語信息處理詞彙 01 部分：基本術語》)
原本指書寫漢語的符號；在本參考指引內特指統一的中日韓文字及《香港增補字符集 — 2001》內的字。
- 3.3 簡化字 Simplified Chinese Character (GB 12200.1-90 《漢語信息處理詞彙 01 部分：基本術語》)
特指內地於1986年公布的《簡化字總表》（見上文第2.2.10段）所界定的字。
- 3.4 UCS-2 編碼 UCS-2 (GB 13000.1-93 《信息技術 通用多八位編碼字符集 (UCS)》)

²台灣工業標準“大五碼”之正式文獻

- 指 ISO/IEC 10646 國際標準中以雙八位定義的編碼系統。UCS-2 編碼以十六進制表示。
- 3.5 **統一碼 Unicode (The Unicode Standard)**
統一碼是一個由業界制定的以雙八位定義的編碼體系。其編碼與 UCS-2 的編碼相同，並包括系統實現的指引。
- 3.6 **原字集 CJK Ideograph Source (ISO/IEC 10646 Annex S 《信息技術 通用多八位編碼字符集 (UCS)》附件 S)**
ISO/IEC 10646 國際標準基本多文種平面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文字標準後得出的。這些原有的電腦文字標準或規範稱為原字集。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)。
- 3.7 **認同規則 Unification Procedure (ISO/IEC 10646 Annex S 《信息技術 通用多八位編碼字符集 (UCS)》附件 S)**
指 ISO/IEC 10646 國際標準中使用的一套漢字統一整合規則³。根據認同規則被認同的漢字不會分別編碼，而在不違反認同規則的原則下，ISO/IEC 10646 國際標準不限制某一碼位漢字的具體字形。
- 3.8 **原字集分別編碼原則 Source Separation Rule (ISO/IEC 10646 Annex S 《信息技術 通用多八位編碼字符集 (UCS)》附件 S)**
凡在原字集中具有各自代碼的異體字不受認同規則限制，在 ISO/IEC 10646 國際標準中予以保留並分別編碼，以保證 ISO/IEC 10646 與原字集之間的對映關係不受影響。
- 3.9 **非同源漢字 Non-cognate Character (ISO/IEC 10646 Annex S 《信息技術 通用多八位編碼字符集 (UCS)》附件 S)**
ISO/IEC 10646 國際標準中之術語。指字形近似但無字源關係的，或意義、用法不同的漢字，因此分別編碼。例如：“士”和“土”。有些國家和地區認為是同源的漢字，在其他國家和地區可能屬於非同源漢字。例如：“瞳” (UCS-2 編碼為“6723”) 和“瞳” (UCS-2 編碼為“81A7”)。ISO/IEC 10646 國際標準把一些容易混淆的非同源漢字載列於附件 S。
- 3.10 **宋體 Song Style (GB/T 12200.2—94 《漢語信息處理詞彙 02 部分：漢語和漢字》)**
現代最通行的漢字印刷體，由宋代刻版所用字形發展而來。此文件中的宋體字形則採用與《香港電腦漢字楷體字形參考指引》內楷體字形一致的原則。
- 3.11 **筆畫 Stroke (GB/T 12200.2—94 《漢語信息處理詞彙 02 部分：漢語和漢字》)**
構成漢字字形的最小連筆單位。
- 3.12 **筆形 Stroke Feature (GF 3001-1997 《國家語言文字工作委員會 語言文字規範 信息處理用 GB13000.1 字符集漢字部件規範》)**
筆畫的具體書寫形狀。漢字的筆形依據筆勢和走向可以分為數十種，其基本的類別為

³ 認同規則原文見 ISO/IEC 10646 Annex S (Informative) Procedure for the Unification and Arrangement of CJK Ideographs.

“一” (橫)、 “丨” (豎)、 “丿” (撇)、 “丶” (點) (包括“丶” (捺)) 和 “㇇” (折)。此文件中的筆形是楷體筆形概念的延伸。

3.13 **筆順** Stroke Order (GF 3001-1997《國家語言文字工作委員會 語言文字規範 信息處理用 GB13000.1 字符集漢字部件規範》)

書寫每個漢字時筆畫的次序和方向。

3.14 **筆數** Stroke Count (GB/T 12200.2—94《漢語信息處理詞彙 02 部分：漢語和漢字》) 構成一個漢字或漢字部件的筆畫數目。

3.15 **漢字部件** Chinese Character Component (GB/T 12200.2—94《漢語信息處理詞彙 02 部分：漢語和漢字》)

由筆畫組成的具有組配漢字功能的構字單位。簡稱“部件”。

例如：“木、心、口、也”。漢字的偏旁部首是部件的一種。

3.16 **基礎部件** Basic Component (GF 3001-1997《國家語言文字工作委員會 語言文字規範 信息處理用 GB13000.1 字符集漢字部件規範》)

最小的不再拆分的部件稱為“基礎部件”。基礎部件處於漢字結構的最低層，又稱末級部件。例如：“男”中的“田”、“力”。

3.17 **主形部件和變體部件** Principal Component vs. Variant Component

某些部件的形狀會受到位置影響而變化。該等部件的原形稱為主形部件；受到位置影響而出現的變體稱為變體部件。例如：“又”(見於“設”字)與“𠄎”(見於“燦”字)兩者之中，“又”為主形部件，“𠄎”為變體部件。主形部件具有代表和稱說本組部件的作用。

3.18 **相關部件** Related Component

有字源關係的兩個或多個部件。例如：“手”和“扌”。

3.19 **形似部件** Similar Component

形狀相似，但無字源關係或主形與變體關係的兩個或多個部件。例如：“土”和“土”。

3.20 **部件拆分** Component Decomposition (GF3001-1997《國家語言文字工作委員會 語言文字規範 信息處理用 GB13000.1 字符集漢字部件規範》)

將漢字拆分為部件稱部件拆分。

3.21 **結構理據** Structure Origin (GF3001-1997《國家語言文字工作委員會 語言文字規範 信息處理用 GB13000.1 字符集漢字部件規範》)

根據字源或參考字源，從漢字的部件組合中分析出的造字意圖，稱結構理據。

3.22 **有理據拆分和無理據拆分** Etymological Decomposition vs. Ad-hoc Decomposition (GF3001-1997《國家語言文字工作委員會 語言文字規範 信息處理用 GB13000.1 字符集漢字部件規範》)

根據漢字結構理據所進行的部件拆分，稱有理據拆分；當無法分析理據或理據與原始形成的字形發生矛盾時，依照字形所進行的部件拆分，稱無理據拆分。

4. 部件拆分原則

本指引對適用範圍內的漢字進行部件拆分，拆分原則以《國家語言文字工作委員會 語言文字規範 信息處理用 GB13000.1 字符集漢字部件規範》第 5 段《漢字基礎部件表》使用規則所載的部件拆分原則為基礎⁴。具體的拆分原則如下：

4.1 沿用《信息處理用 GB13000.1 字符集漢字部件規範》所載部件的原則：

本指引在拆分部件時盡量沿用《信息處理用 GB13000.1 字符集漢字部件規範》已載部件。

4.2 不違背形體的原則：

4.2.1 交重不拆。例如：“戈”不再拆分；

4.2.2 相離可拆。例如：“骨”拆分成“冎”和“冎”；

4.2.3 相接可拆。例如：“藏”中的“𠂔”可拆分成“冎”和“戈”；

4.3 遵照結構規律的原則：

部件組成合乎漢字的結構理據，進行有理據拆分，例如：“骨”拆分成“冎”和“冎”。無法分析理據或理據與字形出現矛盾時，依形進行無理據拆分。例如：“丟”字原由“一”和“去”組成，但由於“王”與“厶”已包括在《信息處理用 GB13000.1 字符集漢字部件規範》內，因此“丟”以無理據拆分成“王”與“厶”。

4.4 合理控制下限的原則：

4.4.1 兩個以上的相離或相接部分，若其中一個無構成其他字的能力，則不再拆分。如其中一部分屬部首則可再拆分。例如：“廡”字拆分成“厂”、“𠂔”和“彡”，由於“𠂔”不獨立構字，因此“𠂔”不會進一步拆分。

4.4.2 兩個相離的對稱部分，不獨立構字，不再拆分。例如：“門”。

4.5 依筆形區分部件的原則：

4.5.1 主形部件和變體部件分別列出。例如：“人”和“亻”。

4.5.2 筆形完全相同的不同源部件，當作同一部件。例如：“陽”字和“都”字中的部件“阝”並不同源（部件位於左邊屬“阜”，部件位於右邊屬“邑”），因筆形完全相同，當作同一部件。

4.5.3 不同源的部件如按香港慣常寫法筆形完全相同，當作同一部件。例如：“月”和“月”為不同源的部件，但這兩個部件出現在字的下方時（例如“肩”和“有”），都寫作“冎”，因此都當作部件“冎”。）

4.6 綜合衡量的原則：

以上原則無法兼顧時，酌情處理。

5. 《香港電腦漢字宋體(印刷體)字形基礎部件表》

本表列載的基礎部件合共 644 個，是以《香港電腦漢字楷體字形參考指引》內的楷體

⁴由於本指引的適用範圍不包括 ISO/IEC 10646 國際標準中的內地規範簡化字及僅在日本、韓國、越南、新加坡使用的漢字，因此拆分出來的部件不包括這批漢字獨有的部件。

字形為藍本，及對本指引適用範圍內的漢字進行拆分的結果。

5.1 部件排列方式

5.1.1 部件按筆畫數從少到多順序排列。

5.1.2 筆畫數相同，則按起筆“橫豎撇點折”順序排列。

5.2 欄目說明

5.2.1 **序號**：列出部件的排列順序號。

5.2.2 **部件**：列出部件的基本形狀和出處代碼。

出處表示由該部件組成的漢字是否出現在以下四種參考文獻所收的字中。出處代碼以四個數字(ABCD)組成：

A：“大五碼”(見 2.2.13 段)

B：《康熙字典》

C：《漢語大字典》

D：《香港增補字符集 — 2001》

每位數字“0”表示該參考文獻所收的字使用該部件，“1”表示沒有使用。由於本指引適用範圍內的漢字出處不一，出處代碼可提供附加的參考信息。

5.2.3 **筆畫數**：列出部件的實際書寫筆畫數目。

5.2.4 **例字**：列出包含該部件的漢字印刷體(宋體)字形作為示例。

5.2.4.1 例字後面有“……”者，表示沒有窮舉，還有其他字使用該部件。所舉例字盡量選用《常用字字形表(二零零零年修訂本)》所載的字(即香港的常用字)，且可以體現部件在構字時所在的不同位置。

5.2.4.2 例字後面沒有“……”者，表示已經窮舉。

5.2.4.3 每個例字下方第一行列出其 UCS-2 編碼。有“*”者表示該字收錄於《常用字字形表(二零零零年修訂本)》。

5.2.4.4 UCS-2 編碼下方標明該例字的出處代碼，格式與 5.2.2 段所述者相同。

5.2.5 **附註**：與部件相關的補充說明。

例如標示主形部件、變體部件、相關部件和形似部件的序號；按原字集分別編碼原則而分別編碼的漢字；非同源漢字；以及字形相似的漢字。

5.2.6 **GB 組號和 GB 序號**：如部件載列於《信息處理用 GB13000.1 字符集漢字部件規範》，該部件在規範內對應的組號和序號也一併列出，以便檢索。由於《信息處理用 GB13000.1 字符集漢字部件規範》不是字形規範，而且內地的書寫習慣與香港不盡相同，因此該規範所列的部件形狀可能與本指引所列對應的部件形狀略有不同。