

《香港增補字符集—2016》“表意文字描述序列” 編訂原則

背景

- 1 中文界面諮詢委員會工作小組(工作小組)在第十三次會議曾討論《香港增補字符集—2016》“表意文字描述序列”(Ideographic Description Sequence (IDS)) (HKSCS-IDS)初稿，特別是當中的編訂原則。

最新進展

- 2 編訂原則已根據會議討論結果修訂，修訂本見附件。
- 3 我們計劃把編訂原則定稿提交中文界面諮詢委員會，以便在其第二十六次會議上審議。編訂原則一經確認，工作小組便會進一步校對 HKSCS-IDS，以期在 2019 年發表該文件。

徵詢意見

- 4 歡迎工作小組成員就上述“表意文字描述序列”的編訂原則修訂本及工作計劃發表意見。

政府資訊科技總監辦公室
2018 年 11 月

《香港增補字符集—2016》

表意文字描述序列編訂原則

簡介

- 1 本文件旨在為《香港增補字符集—2016》(HKSCS-2016)內四千六百多個漢字編訂表意文字描述序列(Ideographic Description Sequence (IDS)) (HKSCS-IDS)，以便更有效地審議和管理字符集內的漢字。
- 2 IDS 利用 ISO/IEC 10646 國際編碼標準所定義的表意文字描述字符(Ideographic Description Character (IDC))，來標示表意文字的結構，並依次列出組成部分。以“朙”字為例，IDS 是“𠄎木月”，表示字符屬左右並列結構，左方為“木”，右方則為“月”。
- 3 由於一些字符字形複雜，IDS 資料未必能完全代表該字形，查閱時應按需要檢視該字的實際字形，此點務須注意。
- 4 此外，字符按不同方式描述，可得出不同的 IDS。有見及此，本文件採用了若干原則，務求令編訂工作更有系統而準則一致。下文詳述有關原則，並列舉例子(包括但不限於 HKSCS-2016 的字符)，加以說明。

編訂原則

- 5 表意文字小組(Ideographic Rapporteur Group (IRG))曾發出表意文字描述序列拆分指引(IRGN1183: Guidelines on IDS Decomposition)。本文件以此為參考。
- 6 IDS 的作用是描述表意文字，包括其結構和組成部分。一般來說，在編訂 IDS 時，會先把字符按結構拆分，並以 ISO/IEC 10646 國際編碼標準所定義的 IDC 標示相關結構，再依次列出各個結構內的組成部分。因應漢字結構，ISO/IEC 10646:2014 列出 12 個 IDC，即𠄎、𠄐、𠄑、𠄒、𠄓、𠄔、𠄕、𠄖、𠄗、𠄘、𠄙、𠄚。表意文字小組在 2018 年 5 月召開的第五十次會議上通過新增兩個 IDC，即𠄛和𠄜。根據 IDC 的拆分方法，字符可拆分為兩個或三個部分。每個組成部分可以是一個漢字或部件，又或另一組 IDS。下文會列舉例子詳加說明。

- 7 為確保一般中文電腦平台都支援本文件所編訂的 IDS，每個組成部分內的漢字和部件都會採用已獲編配 ISO/IEC 10646 碼位的相關字符，次序如下：ISO/IEC 10646 中日韓表意文字基本區、擴展區 A、擴展區 B、中日韓筆劃，以及與部首相關的字區。

不拆分的部件

- 8 不拆分的部件將載於附表一。附表一會列出若干部首及其變體、數詞和一些常見的部件，這些字符已獲編配 ISO/IEC 10646 碼位。在本文件中，這些字符的 IDS 直接以字符本身標示，而作為其他字符的部件時，也無須另行拆分。附表一的例子如下：

序號	碼位	字形
1	2E80	丶
2	2E84	㇇
3	2E88	㇈

相似部件的代表字形和碼位

- 9 一些部件字形非常相似而各有碼位。為了方便編訂 IDS 和提高檢索效率，本文件為這類部件選定了代表字形和碼位。舉“𠃉”和“𠃊”為例，“𠃉”(U+961D)、“𠃉”(U+2ED6，部件在字符左方)和“𠃊”(U+2ECF，部件在字符右方)字形非常相似，本文件以“𠃉”(U+961D)為代表字形和碼位；而“𠃊”有 U+2ECA 和 U+27FB7 兩個碼位，本文件以 U+27FB7 為代表碼位。
- 10 此外，有些字符作為部件時，其字形有輕微變動，例如“羊”(U+7F8A)末筆變成撇筆“𠃉”(U+2EB6)、“牛”(U+725B)末筆變成提筆“𠃉”(U+725C)、“子”(U+5B50)末筆變成提筆“子”(U+5B51)。就這類部件而言，在編訂 IDS 時，本文件會採用原來字符為代表字形。例子如下：

字符	IDS	備註
羶	𠄎羊亶	不採用“𠄎”
犷	𠄎牛宣	不採用“牛”
孫	𠄎子系	不採用“子”

不宜混淆的部件

11 本文件以《香港電腦漢字參考字形》所載的字形為依據。某些部件雖然字形近似，但會分別用以編訂 IDS，而不會以任何一個字形為代表字形，例如“月”(U+6708)、“月”(U+2EBC)、“月”(U+2E9D)和“月”(U+5183)等，不宜混淆。

按筆劃標示的部件

12 按筆劃標示的部件將載於附表二。較難拆分的部件，會以橫(H)、豎(S)、撇(P)、捺(N)、點(D)、折(Z)、彎／彎鉤(W)標示部件的首尾兩筆，再列明筆劃數目，例如字符“𠄎”的上方部件，首尾兩筆為撇(P)和折(Z)，筆劃數目為 7，因此該部件以 PZ7 標示。附表二的例子如下：

筆劃標示	編碼	字形	IDS
PH17	F907	龜	PH17
PZ7	2048E	𠄎	𠄎PZ7 儿

字符結構示例說明

13 下文按字符結構，即並列結構(𠄎、𠄎、𠄎、𠄎)、包圍結構(𠄎、𠄎、𠄎、𠄎、𠄎、𠄎、𠄎、𠄎、𠄎)和重疊結構(𠄎)，列舉例子加以說明。

並列結構(𠄎、𠄏、𠄐、𠄑)

14 凡字符屬上下並列結構，IDC 以先 𠄎 後 𠄐 的方式排列，並由上而下順序描述各組成部件：

組成部分數目	描述方式	例子
2	𠄎 AB	兒：𠄎 白儿
3	𠄐 ABC	賣：𠄐 士四貝
4	𠄎 𠄐 ABCD	會：𠄎 𠄐 人一由曰
5	𠄐 𠄐 ABCDE	𠄐：𠄐 𠄐 一田一田一

15 屬左右並排結構的字符，IDC 按同一方式排列，並由左至右順序描述各組成部件。

16 如字符由四個部分組成，而各部分又各據一方，呈“田”字形狀，則按先上下後左右的方式描述。以“𠄒”字為例，IDS 為 𠄎 𠄐 土土 𠄑 坐坐。

包圍結構(𠄓、𠄔、𠄕、𠄖、𠄗、𠄘、𠄙、𠄚、𠄛)

17 凡字符屬包圍結構，即其中一個組成部件或 IDS 有超過一面為另一部件或 IDS 所攔擋，一律按先外後內的方式描述，例如：

字符	IDS
𠄓	𠄓 冂 𠄎 𠄑 乂 乂 一
𠄔	𠄓 广 𠄎 車 乙

重疊結構(𠄜)

18 凡字符由兩個重疊的部分組成，一律按(𠄜 AB)的方式描述，而組成部件則按碼位排列，例如：

字符	IDS	部件的碼位
串	𠄜 吕	：4E28；吕：5415
隶	𠄜 彡 水	彡：2E95；水：6C3A
夾	𠄜 从 大	从：4ECE；大：5927

表意文字描述序列總表

19 《香港增補字符集—2016》表意文字描述序列總表將載於附表三，詳列字符集內四千六百多個漢字的 IDS，以供參閱。附表三的例子如下：

碼位	字形	表意文字描述序列 (IDS)
74E7	𡗗	𠃉瓦十
8FCC	𡗘	𠃉𠃊月