

中文界面諮詢委員會工作小組 工作進展報告

目的

1. 本文件旨在簡述中文界面諮詢委員會工作小組(工作小組)於 2018 年的工作進展。

工作小組會議

2. 自 2018 年 1 月 11 日中文界面諮詢委員會(中諮會)舉行第二十五次會議後，轄下的工作小組召開了三次會議，日期如下：

2018 年 3 月 16 日：第十二次會議

2018 年 7 月 13 日：第十三次會議

2018 年 11 月 30 日：第十四次會議

審議《香港增補字符集》字符增收申請

3. 在 2018 年召開的三次會議上，工作小組共審議了 125 個字符的增收申請。經詳細審閱後，工作小組建議增收一個字符，其餘無須增收。該字符已納入 ISO/IEC 10646 國際編碼標準字集，建議新增 H-列字源資料，編碼格式為 HD-5C83，詳情如下：

字符	ISO/IEC 10646 碼位	H-列字源資料
𠂇	5C83	HD-5C83

推廣共通中文界面的工作

4. 應表意文字小組(Ideographic Rapporteur Group (IRG))的要求，香港特別行政區政府檢視了多個版本的 *IRG Working Set*，並經由秘書處把檢視結果提交表意文字小組考慮。這些版本包括已在 2018 年 5 月表意文字小組第五十次會議完成審議的 *IRG Working Set 2015*

Version 5.0 和 *IRG Working Set 2017 Version 1.1*，以及已在 2018 年 10 月第五十一次會議完成審議的 *IRG Working Set 2015 Version 6.0* 和 *IRG Working Set 2017 Version 2.0*。

5. 為協助審議和管理字符，以及促進中文輸入軟件和字型軟件的開發，工作小組正嘗試為《香港增補字符集—2016》內的漢字編訂表意文字描述序列 (Ideographic Description Sequence (IDS))，並已訂立若干編訂原則(詳見附件)，方便日後檢索 IDS 之用。《〈香港增補字符集—2016〉表意文字描述序列》(HKSCS-IDS) 的初稿已經完成，並正作進一步校對，預計可在本年內定稿。
6. 工作小組在討論推廣共通中文界面的題材後，共製作了三部動畫，幫助公眾認識 ISO/IEC 10646 國際編碼標準，以及推廣使用獲編配 ISO/IEC 10646 碼位的字符作中文電子通訊。該三部動畫已先後在 2018 年 10 月至 11 月推出，主題如下：
 - “等同的表意字符”
 - “什麼是‘表意文字描述序列’？”
 - “無法在電腦輸入的字符，是否就要自行造字？”

一般查詢

7. 在支援共通中文界面方面，在 2018 年 1 月至 2018 年 12 月期間，秘書處共處理了 23 宗由政府部門提出的查詢，以及 30 宗由公眾人士提出的查詢。

中文界面諮詢委員會秘書處
2019 年 1 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))，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“朧”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、𠄏、𠄐、𠄑、𠄒、𠄓、𠄔、𠄕、𠄖、𠄗、𠄘、𠄙。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 𠄚和 𠄛。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(基本區)、擴展區 A、擴展區 B、中日韓筆劃，以及與部首相關的子區。
8. 視乎字符拆分方式，一個字或會有多個可行的 IDS。原則是選取組成部件最少者作為 IDS。
9. 如 IDS 組成部分有屬於獲編配於 2017 年 12 月出版的 ISO/IEC 10646:2017 擴展區 C 至 F 的字符，附註欄會多提供一組 IDS，這些在擴展區 C 至 F 的字符會直接以碼位標示。
10. 以《香港增補字符集》字符“𠵽”(U+25C01)為例，該字的 IDS 是 𠵽^ㄅ 𠵽^ㄆ 方生，IDS 組成部分“^ㄅ”(U+25AD7)屬擴展區 B 的字符，“方”(U+65B9)和“生”(U+751F)則是基本區字符。由於“𠵽”字下方部件“𠵽” (U+2D92F)屬擴展區 F 的字符，附註欄會多提供一組 IDS： 𠵽^ㄅ U+2D92F。

碼位	字形	表意文字描述序列 (IDS)	附註
25C01	𠵽	𠵽 ^ㄅ 𠵽 ^ㄆ 方生	𠵽 ^ㄅ U+2D92F

不拆分的部件

11. 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	乚
273	9580	門

相似部件的代表字形和碼位

12. 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“𠃉”(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。
13. 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“𠃉”(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
羴	𠃉羊亼	不採用“𠃉”
犒	𠃉牛亼	不採用“𠃉”
孫	𠃉子系	不採用“𠃉”

不宜混淆的部件

14. 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“冂”(U+5183)等，不宜混淆。

按筆劃標示的部件

15. 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“𠃉”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	表意文字描述序列 (IDS)
PH17	F907	龜	PH17
PZ7	2048E	皃	皃 PZ7 儿

重疊結構(𠄎)

16. 凡字符由兩個重疊的部分組成，一律按(𠄎 AB)的方式描述，而組成部件按筆順排列，例子如下：

碼位	字形	IDS
4E24	兩	𠄎一 𠄎冂从
4E32	串	𠄎吕
26612	聿	𠄎 聿一

表意文字描述序列總表

17. 《〈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)	附註
74E7	𠄎	𠄎瓦十	
8FCC	𠄎	𠄎辵月	
25C01	𠄎	𠄎 ^々 𠄎方生	𠄎 ^々 U+2D92F