

## 討論編訂表意文字描述序列的工作

### 目的

- 1 工作小組擬為《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) (HKSCS-IDS)，以方便字符的審議和管理，以及促進中文輸入和字形相關軟件的開發。

### 表意文字描述序列簡介

- 2 表意文字描述序列就是把字符分拆成較簡單的部件，以 ISO/IEC 10646 國際編碼標準所定義的 12 個表意文字描述字符 (Ideographic Description Characters, IDC) (𠄎 𠄐 𠄑 𠄒 𠄓 𠄔 𠄕 𠄖 𠄗 𠄘 𠄙 𠄚) 來表示部件的分拆方法，例如上下(𠄎)、左右(𠄐)、包圍(𠄑)和重疊(𠄒)等，並依次列成 IDS。IDS 的例子如下：

《香港增補字符集—2016》 字符	表意文字描述序列 (IDS)
膾	𠄐月 𠄎內口
𠄎	𠄑光日

### 編訂表意文字描述序列的原則和方法

- 3 字符是根據《香港電腦漢字參考字形》所載字形來進行分拆的。
- 4 編訂 IDS 的原則和程序見於表意文字小組文件 IRGN1183 IDS decomposition principles<sup>1</sup>。表意文字小組會在 2018 年 5 月舉行的第五十次會議討論新增四個 IDC 的事宜。這四個 IDC 是包圍(𠄑, 𠄒), 左右反映(𠄐)

<sup>1</sup> <http://appsrv.cse.cuhk.edu.hk/~irg/irg/irg25/IRG25.htm> Document N1183

和 180 度旋轉 (☐)。若會議通過採納，我們建議相應採用新增的 IDC，以便利 HKSCS-IDS 的編訂工作，用以分拆類似以下例子的《香港增補字符集—2016》字符：

字符	IDS
氷	☐水、

- 5 編訂 IDS 的概略流程見附圖，該流程以 IRGN1183 訂定的程序為藍本，並加入下列建議，以期令 IDS 的編訂工作更為一致，並能利用軟件檢測，方便日後進行討論和檢索。

[上述表意文字小組文件中用以編訂 IDS 的漢字並不限於 HKSCS 漢字]

- 5.1 初步訂出一些不分拆的字形或部件，在 IDS 欄內直接以該字標示。

不分拆的字形包括大部分部首和相關字形、數詞和一些常見於其他字符的部件。例子如下：

不分拆的字形或部件	不分拆的原因	包含該字形的字
鬥	部首。不分拆為左右兩部分。	鬪：☐鬥☐豆寸
三	數詞。不分拆為 ☐一一一或 ☐一二	弎：☐弋☐三
我	常見於其他字符	俄泐峨哦娥 哦餓義鵝鷺

同一字形有多個編碼(例如 ☐(961D)、☐(2ED6)、☐(2ECF))的情況會由軟件處理，以人手編訂或檢視時無須兼顧。

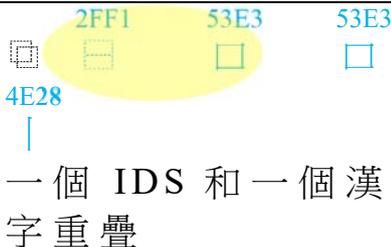
- 5.2 根據 IDC 的分拆方法，一個字形可以分拆成兩個或三個部分，而每個部分可以是一個漢字、部件或另一組 IDS。現時常用的電腦平台支援約七萬多漢字(包括 ISO/IEC 10646 中日韓表意文字區及其擴展區 A 和 B)及一些部首，但並未內置表意文字區擴展區 C 至 F (約一萬八千多漢字)的字型。我們建議編訂 IDS 時首先採用現時常用平台所支援的漢字，這樣做可以令編訂 HKSCS-IDS 的工作更方便。
- 5.3 沒有編碼而結構複雜的字形或部件(例如多個部件重疊)，則以橫(H)、豎(S)、撇(P)、捺(N)、點(D)、折(Z)方式，標示該整個部件的首尾兩筆和總筆劃數，例子如下：

字符	IDS
	SH13
	☐ PZ7 几

待 HKSCS-IDS 的初稿完成後，我們會一併處理這些字形，例如檢視這些字形是否已收納在表意文字區擴展區 C 至 G 內。

表意文字小組文件 (IRGN2225 CJK Supplementary Components for IDS Use)載列一些沒有編碼的字形或部件，可作參考之用。

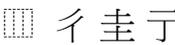
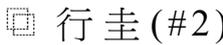
5.4 當字形可分拆成重疊的部件時，**重疊的兩個部分** (可以是漢字、部件或 IDS)以該部分或 IDS 內首個 IDC 的 ISO/IEC 10646 碼位由小至大列出，例子如下：

字符	列出的IDS	不列出的IDS
来	 兩個漢字重疊	7C73 4E00 
串	 一個IDS和一個漢字重疊	4E28 2FF1 53E3 53E3 

#1 字符上的數字(ISO/IEC 10646 碼位)只在本文件中用作說明排序之用。

由於排序將以電腦軟件處理，因此以人手編訂或檢視包含的IDS時無須兼顧排序。

5.5 若字符結構可分拆成三部分(例如“左中右”)，除了會優先檢視 IRGN1183 所列明“中右”兩個部件外，也一併檢視“左中”和“左右”兩個部件能否組成一字，如有則一併列出。例子如下：

字符	IDS
街	2FF2  2FFB 
澈	2FF0  2FF2  2FFB 

#2 若“左中右”結構的左右兩個部件能組成一字，IDC 採用重疊  表示。可分拆成“上中下”三個部件的字符亦以同樣方式處理。

5.6 當一個字符可分拆成多於一組 IDS 時，建議根據碼位由小至大列出。例子如下：

字符	IDS
串	<p>2FFB 2FF1 53E3 53E3 4E28    □ □  </p> <p>2FFB 4E2D 53E3   中 □</p> <p>如兩個 IDS 的首個字相同，則按隨後的字排列。由於碼位 2FF1 小於 4E2D，所以先列出前者的 IDS。</p>
来	<p>2FFB 4E00 7C73   米</p> <p>2FFB 4E37 572A   未</p>

同樣，以人手編訂或檢視多個 IDS 時，無須兼顧排序。

5.7 擬訂 HKSCS-IDS 的工作包括擬備以下文件：

- 編訂表意文字描述序列的原則及方法
- 《香港增補字符集—2016》內漢字表意文字描述序列表  
(序列表的附表會列出 IDS 欄內不再分拆的字形或部件)

## 徵詢意見

6 歡迎中諮會工作小組成員就上文第 3 至 5 段所載建議發表意見。

政府資訊科技總監辦公室

2018 年 3 月

附圖：編訂“表意文字描述序列”概略流程示意圖

