

## 討論如何落實新字符集的編製工作

### 背景

《香港增補字符集》自 1999 年初版，其後三度更新，出版《香港增補字符集－2001》、《香港增補字符集－2004》和《香港增補字符集－2008》。三個版本分別增收 116、123 和 68 個字符。

2. 支援 ISO/IEC 10646:2003 的平台 2009 年開始普及，內置漢字逾 70 000 個。2008 年 4 月至 2015 年 3 月，中文界面諮詢委員會（中諮會）接獲逾 700 個《香港增補字符集》字符增收申請；經審核後，確定所有字符已納入 ISO/IEC 10646 國際編碼標準字集，無須增收，可見 ISO/IEC 10646 國際編碼標準大體上能滿足本地中文電子通訊的需要。

3. ISO/IEC 10646 國際編碼標準最新版本 ISO/IEC 10646:2014 包含漢字逾 80 000 個，它整合了不同地區的漢字及提供電腦平台整體而全面的支援。然而，在電腦的實際使用中，支援全版的 ISO/IEC 10646:2014 字集，不僅會增加開發成本及用戶購買產品的費用，降低電腦的效率，而且在日常生活的使用中往往沒有必要。所以在實際應用上，為支援平台需要的字體或輸入法開發商會因應個別地區的用字要求，選取適量字符開發支援產品。再者，個別地區的使用者亦希望能夠優先選取在該地區較常用的字符及字形。

4. 目前開發者按產品規格和客戶要求而選取適量字符。因此在 ISO/IEC 10646 國際編碼標準字集內註明哪些字符是香港用字及加入香港字形，可減少開發成本和時間，有利資訊行業開發更適合香港的產品，藉此進一步完善本地中文電子通訊的需要。

5. 中文資訊科技工作小組和中文電腦用字工作小組分別在 2015 年 4 月 29 日和 5 月 14 日的會議上，原則上贊成編製《香港字符集》（暫定名稱），作為 ISO/IEC 10646《通用字符集》內一個完整命名字符集，這個新字符集即是 ISO/IEC 10646 的一個子集，定義香港電腦平台所慣常使用的字符。為有效推展新字符集的編製和推廣工作，現提出多項考慮要點，以供討論。

6. 新字符集包括了《香港增補字符集—2008》所有字符，但增補是對大五碼而言的，也就是說，香港的慣常用字，遠不止《香港增補字符集—2008》內包括的字符。作為香港特區對所用字符的定義，以《香港字符集》命名比較適宜，但有可能被狹義理解為收錄香港特有字的字集。選擇適當名稱，減少誤解，或有助向公眾推廣新字符集。附件一列出一些中英名稱建議，以供參考及討論。

7. 當年編製《香港增補字符集》並制訂增收字符機制的目的，是解決中文電子通訊的主要問題，包括編碼標準不一、部分中文電腦字集的字符不齊全。在中諮會委員和業界的支持下，ISO/IEC 10646 國際編碼標準和《香港增補字符集》及其增收字符機制已得到廣泛認同和採納。

8. 新字符集的增收字符機制除了應保留向 ISO 提交新字（縱向擴展）的功能外，也應讓香港能夠因應需要，為 ISO/IEC 10646 國際編碼標準字集內適用於香港的漢字，加入香港字形（橫向擴展）。要新字符集能持續和更有效地滿足香港的中文電子通訊的需要，有賴市民，用戶和軟件開發商透過增收字符機制提交用字需求。附件二初步提出新字符集及增收字符機制的編製方案，以供討論。

9. 新字符集初稿預計收錄字符逾 18 000 個：5 009 個《香港增補字符集—2008》字符和逾 13 000 個大五碼字符。為求新字符集初稿收錄的字符更完備，可考慮根據新制訂的字符增收原則和程序增收新字，尤其是橫向擴展的字符。此舉可讓公眾明白橫向擴展的具體運作。附件三列出新字符集初稿可考慮增收的字符，以供討論。

10. 中諮會為《香港增補字符集－2008》編製的參考資料，例如部首、倉頡碼、粵語拼音、字形等，有助業界開發支援本地要求的產品。在新字符集出版後，這些資料亦應隨之更新。Unicode Consortium 的 Unihan database 已包括大部分大五碼字符的參考資料，現建議編選 Unihan database 有關參考資料，分批提交工作小組審議。

## 建議

11. 秘書處會收集工作小組成員對上述建議的意見，整理後提交中諮會審議。

## 徵詢意見

12. 請工作小組成員發表意見。

2015年8月

### 新字符集命名建議

技術上，新字符集是 ISO/IEC 10646 的一個子集，所以建議當中也包括子集的命名作參考。

	中英名稱建議	備註
1 a	《香港字符集/子集》 Hong Kong Character Set/Subset (HKCS)	中文名稱可能被解讀為香港的中文文字集。
1 b	《香港特區字符集/子集》 The Character Set/Subset for the Hong Kong Special Administrative Region (HKSARCS)	英文縮寫 HKCS 略短，不容易明白。  英文縮寫 HKCCS 與 HKSCS 相近，容易混淆。
1 c	《香港中文字符集/子集》 Hong Kong Chinese Character Set/Subset (HKCCS)	
2 a	《香港共通/通用字符集/子集》 Hong Kong Common Character Set/Subset (HKCCS)	能夠表達中文電子通訊的需要，但中英名稱略長。
2 b	《香港中文界面字符集/子集》 Hong Kong Chinese Language Interface Character Set/Subset (HKCLICS)	英文縮寫 HKCCS 與 HKSCS 相近，容易混淆。
2 c	《共通中文界面字符集/子集》 Common Chinese Language Interface Character Set/Subset (CCLICS)	

### 新字符集編製方案及程序

為了更有效推展新字符集的編製工作，現建議先制訂一套同時適用於縱向及橫向擴展的字符增收原則和程序，然後據此編製新字符集初稿。新字符集初稿完成後，再連同字符增收原則和程序，進行公眾諮詢，蒐集業界和公眾對新字符集整體編製工作的意見。

編製新字符集的主要工作及程序建議如下：

	主要工作及程序	負責單位
1	制訂新字符集字符增收原則和程序。	工作小組
2	根據新字符集字符增收原則和程序，審議 ISO/IEC 10646 國際編碼標準字集的橫向擴展建議，為字集內適用於香港的漢字加入香港字形。	工作小組
3	編製新字符集初稿，收錄大五碼字符、《香港增補字符集－2008》字符和橫向擴展字符。	工作小組
4	審議新字符集初稿及增收原則和程序。	中諮會
5	就新字符集初稿及增收原則和程序進行公眾諮詢。	秘書處
6	正式出版新字符集。	秘書處
7	向 ISO 轄下表意文字小組提交橫向擴展字符，以期納入 ISO/IEC 10646 國際編碼標準日後的版本中。	秘書處

新字符集增收字符建議

在擬訂新字符集初稿時，可以根據新字符集字符增收原則和程序，審議增收以下字符：

- 須分別編碼以保留大五碼字形與香港字形的字符（22個）
- 已收錄於 ISO/IEC 10646 國際編碼標準字集但不屬大五碼或《香港增補字符集》的字符申請（約 240個）

建議除了大五碼字符和《香港增補字符集—2008》字符外，其他字符都應經工作小組審核，確保新字符集的增收字符準則一致。