中文界面諮詢委員會
停止為《香港增補字符集》編配 ISO/IEC 10646
國際編碼標準私人使用區編碼

背景

　　為了解決在中文通訊和交換資料的時候用戶自行定義造字引起資料失誤的問題，特區政府與中文界面諮詢委員會（中諮會）合作，編訂《香港增補字符集》。

2. 　　《香港增補字符集》有兩套編碼方案，一套適用於 ISO 10646 國際編碼標準平台，另一套適用於大五碼系統。

3. 　　未包含在 ISO 10646 國際編碼標準內的《香港增補字符集》字符，會編入相應 ISO 10646 國際編碼標準版本內的私人使用區(PUA)內。在大五碼系統內，《香港增補字符集》的字符則編入使用者造字區(UDA)內。兩套編碼互相配對，以支援在這兩種平台上進行資料互換。

4. 　　由 2008 年 3 月 31 日起，新增收的《香港增補字符集》字符已不獲編配大五碼使用者造字區編碼。另外，在 2009 年 3 月底，所有《香港增補字符集》字符已獲 ISO 10646 國際編碼標準編碼。

建議停止編配 ISO 私人用字區編碼

5. 　　有見及此，"中文資訊科技工作小組"召集人陸勤教授提出建議，停止為新增收的《香港增補字符集》字符編配 ISO 私人使用區編碼 (詳見附件一)。

6. 　　政府資訊科技總監辦公室亦草擬了 "於香港特區採用 ISO/IEC 10646 標準" 文件闡明特區政府的立場 (見附件二)

7. 　　請各委員就附件一之建議發表意見、討論，並定案確認，讓秘書處跟進執行。

第 2 頁

中文界面諮詢委員會秘書處
2009 年 **9** 月

The Proposal to Stop Encoding of Characters in ISO PUA

By Lu Qin
For discussion by the IT subgroup of CLIAC
June 16, 2009

Background: Restricted by the coding structure of Big5, HKSCS had to assign code points for its collection in the User Defined Areas(UDAs) which are in principle not exchangeable. Yet, this is manageable under locale dependent code page operating mode (HKSCS-Big5) and technical viable to support in the old code page based implementations. To help the migration to ISO 10646, HKSCS also provided mapping of HKSCS-Big5 to ISO 10646 codepoints (HKSCS-UCode).

In the earlier stages of HKSCS-Big5 development, some characters were not encoded in ISO 10646. To support round trip conversion, the unencoded characters of HKSCS in ISO 10646 are mapped into the Private User Area(PUA) purely for round trip conversion back to HKSCS-Big5 environment.

At present, HKSCS has stopped the encoding of characters in Big5. This is a very important strategic step by the Hong Kong Government to move Hong Kong's computer systems from a locale dependent platform to a more universally used platform which supports multilingual scripts. As at the end of March 2009, all HKSCS characters are encoded in ISO 10646. The fact that HKSCS no longer encode its characters in UDAs warrants a serious consideration to stop coding HKSCS characters temporarily in the PUA of ISO 10646.

Summary of Proposal: Revise the current practice for the temporary assignment of newly accepted HKSCS characters in the PUA. Instead, propose them directly to ISO 10646 Working Group for inclusion and code point assignment.

Rationales:
1. **Principle contradiction**. The mapping of HKSCS to PUA has some intrinsic problems in ISO 10646 platforms. The PUA area is designated for private use and is thus not for exchange. In Principle, it is a contradiction for HKSCS to assign characters in PUA as PUA assignment is not part of standardization in ISO 10646 platforms.
2. **Technical Difficulty Due to the Nature of PUA**: Being a PUA, system platforms ignore the data in PUA as it is by definition private and can be either, user, location, or application dependent. Without knowing the user, the location, or the application, management of PUA by systems is not possible. It is very different from the traditional code page based systems where code page under a particular locale has its locale indicator in the system so that conversions can be done. But most of the locale dependent coding standards, such as Big5, are frozen in the sense that they are no longer being extended. Thus, providing backward compatibility support does not require continuous change because the mapping table is fixed. In fact, the support for code pages will eventually fade out.
3. **Technical Difficulty in Seeking Vendor Support**: During the development of HKSCS, 3 versions were produced. If software vendors are to support them, 3 different versions of backward compatibility support must be provided. This cost will continue to rise if HKSCS is to be extended indefinitely in all new platforms

which is a technical burden on the part of vendors, and a cost that most of the software users will need to bear in one form or the other.

Exactly because of this problem, HKSCS mappings to ISO 10646 are supported only in very limited platforms(Microsoft platform is such an example). Other platforms like Apple machines and many Unix platforms simply choose to ignore HKSCS. This disparity will continue in the future, which means that the exchange of data from HKSCS supporting platforms and non-HKSCS supporting platforms cannot guarantee round-trip conversion.

4. **Opportune Time to Change**.   HKSCS has passed the crossroad of code paged implementations and turned to the direction of a more comprehensive, technically sound and globalized encoding. All our HKSCS characters are now accepted by ISO 10646, and HKSCS has also stopped adding more characters in HKSCS-Big5. What this means is that the code paged based HKSCS-Big5 is frozen which gives a good opportunity for vendors of every platform to support HKSCS 2008 once and for all as they know there is no more extensions of HKSCS-Big5. In the mean time, HKSCS-Big5 mapping to ISO 10646 are readily available so that they do not need to deal with PUA which many Unicode vendors have no way to handling them. All new characters of HKSCS can be directly included in ISO 10646.

5. **Inclusion time by ISO 10646 is reasonably short.** From the current practice and our successful inclusion of the last 7 HKSCS character in ISO 10646, we know that the inclusion of new characters can be done in a reasonable time frame of one year which is acceptable in any inclusion of characters in a standard.

Proposed course of actions:
1. Revision of acceptance of character procedure:
   When a character is accepted against ISO 10646 standard, no temporary PUA codepoint will be assigned. The proposer will be informed that the character is accepted under the condition that it will be accepted by ISO 10646.

2. Submit conditionally accepted characters in a proposal to ISO/IEC JTC1/SC2/WG2 and its sub-group IRG for inclusion in ISO 10646.

3. If preliminary review by IRG is to accept the character, such character can be considered accepted pending codepoint assignment.

4. Once codepoint assignment is confirmed, the proposer can be informed of such and its inclusion in HKSCS can be confirmed.

Request for discussion:
1. Input on this proposal and feasibility
2. Revision to current procedures of character acceptance(basically the assignment of code points).
3. Other technical difficulties which we may encounter because of this change.
4. Procedures/process that are required to smooth the transition.

於香港特區採用 ISO/IEC 10646 標準

香港特別行政區政府大力支持採納國際標準化組織(ISO)／國際電工委員會(IEC) 10646 標準,作爲處理中文字符的關鍵標準。政府資訊科技總監辦公室於一九九九年成立有廣泛業界參與的"中文界面諮詢委員會"(下稱"中諮會"),推動各界共同建立香港的中文界面標準,促進電子通訊。中諮會的重要功能之一,是就制定《香港增補字符集》以納入 ISO/IEC 10646 成爲標準字符,提出建議。

2. 至二零零九年八月爲止,已公布的《香港增補字符集》全數 5 004 個字符已成爲 ISO/IEC 10646 標準字符。預計另外五個新的字符也會於二零零九年納入 ISO/IEC 10646:2003 年版的第六修訂版。中諮會也提供意見,推廣採用 ISO/IEC 10646 標準來處理香港增補字符。

3. 政府資訊科技總監辦公室認爲香港特區政府與市民大眾均應該能夠採用 ISO/IEC 10646《通用字符集》(universal character set) 之最新標準版本。而這取決於兩個重要因素:(一)資訊科技產品能否支援最新版本的 ISO/IEC 10646 標準;(二)爲確保使用不同軟件產品版本的應用系統可以互用、以及爲確保使用早期產品版本建立的資料可作存取,而所需的適當工具是否有所供應。

4. 由於**標準產品支援**[1]不足,迄今仍有很多用家未能利用最新版本的字符集來處理資訊和交換資料。爲了應付迫切的實際需要,本地的用戶以往依賴"私人使用區"("PUA")中文編碼等特別的軟件功能及相關的工具去處理和交換包含增補字符的資料。長久依賴**非標準的軟件功能**會產生系統保養問題,是**不可取**的。

---

[1]**標準產品支援** 指以 ISO/IEC10646 標準所需版本編碼建立和處理電子紀錄的能力。

5. 政府資訊科技總監辦公室呼籲各主要產品供應商為最新版本的《通用字符集》提供及時的**標準產品支援**，並同時為現有系統提供所需的**互用工具**[2]以保障用戶已作的投資。

政府資訊科技總監辦公室
二零零九年八月

---

[2]**互用工具** 是一些工具，能方便準確讀取以 ISO/IEC10646 及香港增補字符標準任何版本編碼的電子紀錄。互用工具應該可以應用於仍獲市場支援的所有操作系統。