

## 中文界面諮詢委員會

### 中文資訊科技工作小組

#### 進展報告

2012.11.23

#### 1. 主要工作內容

工作小組在中文界面諮詢委員會第十九次會議後，召開了一次會議（第二十八次會議：二零一二年八月三十一日）。主要有三項議題

- 匯報「表意文字小組」第三十八次會議相關事項
- 闡釋 Ideographic Variation Database（表意文字異體數據庫）及相關技術
- 關於改進香港字形指引的討論

#### 2. 「表意文字小組」(IRG) 第三十八次會議相關事項

秘書處報告香港特別行政區跟進IRG相關的編輯工作。秘書處向與會者簡介根據ISO/IEC和IRG的《原則和程序》（“IRG Principles and Procedures”），增收“迫切需要編碼的字符”（“Urgently Needed Character”）的原則和必須編配 Source Reference（“來源參考”）。因為香港如有新字提出，香港需要提出來源並給與來源編碼。陸勤教授建議來源參考的格式定為HC-9999（“HC-”是固定的數值，代表《香港增補字符集》需要編碼的字符；“9999”是變數），由HC-0001開始，數值每次遞增1。使用“HC-”序列的原因是ISO10646/Unicode 已將HKSCS的字符標識為“HA-”並將大五碼內的其他字符標識為“HB-”（因為HKSCS被認為是大五碼的擴充集），因此“HC-”可以較好的標識與前兩者的相關性和不同性。與會者經討論後，一致通過陸勤教授的建議。有關建議經秘書處整理後，將會提交中諮會討論和確認。

#### 3. 闡釋 Ideographic Variation Database（表意文字異體數據庫）及

## 相關技術

陸勤教授向與會者簡介一個利用異體字選擇符 (Variation Selector) 來表述中文異體字的新技術。該技術給過去在漢字認同規則下不能編碼的異體字進行編碼提供了解決方案，通過由已編碼的漢字和異體字選擇符組成的異體字表述串 (ideographic Variation Sequence, IVS) 使得未編碼的異體字可以同已編碼的漢字相關聯並編碼，再經 Unicode 的表意文字異體數據庫進行註冊，使這些異體字得以在 ISO/IEC10646 體系下編碼。每個表意文字字符最多可以關聯 240 個異體字 (“variants”) 字形。

### 4. 關於改進香港字形指引的討論

有成員表示，開發字型軟件時遇到的困難，主要是不能確定字符的字形究竟是依照香港特區的字形，還是台灣的字形？雖然香港特區已編訂一套《中文電腦字形參考指引》，但因為該指引是基於部件的，廠商很難跟從。主要是跟從該指引所製作的字形結果無法進行有效認證。陸勤教授表示，在編訂《中文電腦字形參考指引》時，無法直接對字符的字形進行規範，是因為有技術上和政策上的困難。現在時機較為成熟，並有 IVD 技術，因此建議需要編訂在香港特區常用、但筆劃結構與大五碼標準字符不同的香港特區中文字形指引，供字型軟件發展商參考。與會者經討論後，一致通過陸勤教授的建議。另外，余志明先生表示有興趣設立和註冊一套顯示《四庫全書》字形的 IVD。