

公布《香港增補字符集—2008》
及
字符增收原則的最新修訂

張群顯

(香港理工大學)

中諮會委員，中文用字工作小組召集人

漢字今天的三大體系

	傳統體系	中國內地體系	日本體系
語言	漢語、韓語	漢語	日語
本質	繼承從唐以來的傳統	五六十年代漢字整理、改革結果	日本對中國傳統的修改
使用地域	台灣、香港、澳門、韓國	中國內地、新加坡、馬來西亞、聯合國	日本
字例	發明, 頭髮	发明, 头发	発明, 頭髮

華人的地區性編碼字符集

年分	字集名	地區	字數	字彙性質	漢字體系
1981	國(家)標(準)	內地	6763	現代漢語用字	中國內地體系
1984	Big5/大五(碼)	台灣	13053	現代漢語用字	傳統體系
1999+	香港增補字符集↘	香港	4941+	對大五(碼)字集的增補	三體系混雜

早期 電腦漢字集 特點 1: 區域性

- 區域性
 - ⇒內地/台灣
 - ⇒香港(、澳門)特用字 缺乏支援
 - ⇒港(、澳)特用字 是 漢字世界的孤兒
- 參張群顯 《漢語信息技術與香港語言生活的配合問題》,
- 載 北京 《語文建設》, #109:20-24, 1997.7.

早期 電腦漢字集 特點 2: 時代性

- 時代性
 - ⇒ 限於現代
 - ⇒ 只限於(現代)常用字
 - ⇒ “不常用字” 缺乏支援
- 常用與否：與“區域屬性”相關
 - ⇒ 加深了“香港(、澳門)特用字 缺乏支援”的嚴重性

大五碼字集未收字舉例

未收字	已收字	推測 缺字 原因
綫	線	異體字
綉	繡	異體字
滙	匯	異體字
栢	柏	異體字
峯	峰	異體字
穎	穎	異體字
邨	村	認為是異體字
着	著	認為是異體字
双	雙	簡化字
𪗇、埜		台灣不用

香港特用字 自強 史

年分	方式	字碼環境	主要問題
1992 前	隨機隨“機” 造“字”(電腦 字符)	大五	重複勞動、每 機獨立
1992	ISO10646內59 字	統一 (Unicode)	字數少、字碼 當時非常規平 台
1995	政府通用字庫↘	大五	認受性 不高
1999	香港增補字符集 → ISO10646	大五及統 一	(對字符輸入 支援不足)

ISO10646內 59個香港字

- ISO10646 定案前 最後關頭 學術界的 游說
- 香港語言學學會 臨急受托
- 受限於 當日 香港用字 研究的 不足
- 提交: 二百來字
- “𨮑、邨、埗” 等字不在提交字集內

- 真正需要增收者: 59字

《香港增補字符集》

- <http://www.ogcio.gov.hk/ccli/unicode/hkscs/introduction.html>(政府網址)
- **Hong Kong Supplementary Character Set**
- 簡稱 **HKSCS**
- 是香港特區政府基於繁體中文電腦操作環境中最流行的大五碼之上擴展的字符集標準
- 是現時香港的中文資訊交換內碼標準
- 前身:《政府通用字庫》(**Government Common Character Set**, 簡稱 **GCCS**) ↘

《政府通用字庫》

- 1990年代初期，香港電腦應用逐漸普及。
- 香港政府各部門使用 Big-5 的外字區，自行補上這些字，並在政府內部使用。
- 到1995年，互聯網在香港起步，而政府也推出了自己的網站。
- 各人電腦的中文系統雖然用 Big-5 編碼，但沒有政府用的外字，在瀏覽網頁時不能正確顯示字符；
- 香港政府於是把內部使用的 Big-5 外字集公開，讓各界可以下載安裝這批字，使電腦能顯示正確的字符，名爲「政府通用字庫」。

從 政府通用字庫 到 香港增補字符集

- 政府通用字庫 本來只是香港政府內部統一使用的造字檔，有三千多字。
- 1999年成立 中文界面諮詢委員會/CLIAC
- 同年, 中文界面諮詢委員會對此字集進行研究整理, 增加到四千多字，並以“香港增補字符集”的名稱對外公布。
- 此 字符集 的認受性比 政府通用字庫 的高
- 負責字符集管理的為中文界面諮詢委員會↘

中文界面諮詢委員會

- 簡稱「中諮會」
- 英文名 Chinese Language Interface Advisory Committee (CLIAC)
- 1999年5月由前資訊科技署成立
- 現隸屬「政府資訊科技總監辦公室」(OGCIO)
- 宗旨: 為配合制訂共通中文界面
- 成員來自學術界、語言學界、出版界和資訊科技界等

中文界面諮詢委員會 職能

- 整體職能: 就在本港制訂和推廣共通中文界面的各個方面，向政府提供專業意見，一直致力制訂和推動共通中文界面。
- 中諮會轄下有兩個工作小組:
 - 中文電腦用字工作小組:
審核待增收字符, 納入和編配碼位等工作
 - 中文資訊科技工作小組:
解決技術問題及與國際標準 ISO/IEC 10646
(簡稱 ISO 10646) 接軌等工作

香港增補字符集的擴編

- 過去10年，中諮會制訂和公布了多個版本的《香港增補字符集》，並提交國際標準化組織，全部獲收編入國際編碼標準字集。

版本	字數	發表時間
HKSCS-2004	4,941	2005年5月
HKSCS-2001	4,818	2001年12月
HKSCS-1999	4,702	1999年9月
(GCCS)	3,049	1995年

香港增補字符集-2008

- 5 009個字符
- 至2009年10月，全數字符經已正式納入ISO/IEC 10646:2003 國際編碼標準字集第6修訂版內。
- 對HKSCS-2004新增68字資料見：
<http://www.ogcio.gov.hk/ccli/unicode/hkscs/download/newchar.pdf>

新增字符舉例

大五碼 字 Unicode 公布日期

87D1	𠄎	224BC	2.2008
87D2	𠄏	9FC8	2.2008
87D3	𠄐	224C1	2.2008
87D4	𠄑	224C9	2.2008
87D5	𠄒	224CC	2.2008
87D6	𠄓	9FC9	2.2008

《香港增補字符集》


新增字符 編碼原則 修訂

- 《字符集》字符過去一直是大五碼和Unicode雙重碼位編配
- 《字符集》在大五碼編碼使用者造字區內預留的碼位無法擴充
- 由 **2008年3月底**起，中諮會對《字符集》新增收字符只會提供 **ISO 10646** 國際編碼標準編碼。
- 之前在《字符集》內已獲編配碼位的字符，一概不受影響，可繼續使用；
- 但鼓勵使用者盡可能將應用系統由大五碼編碼過渡至 **ISO 10646** 國際編碼。

“新增字符 編碼原則 修訂” 含意

- 脫離對 大五碼字集 的 依附關係
- 獲收編入 “國際編碼標準字集” 為 新增字符 的實質存在依據
- 須進一步向 “國際編碼標準字集” 收字準則 靠攏

《香港增補字符集》 字符增收原則 修訂

- <http://www.ogcio.gov.hk/ccli/unicode/hkscs/applicn.html>
- 爲配合《字符集》新增字符編碼原則修訂，《字符集》的《字符增收原則》也相應作出修訂。
- “2.1 香港社群用字”下的(b)條：
 - “此前已流通的專有名詞用字（須提供**依據**）；”

 - “**能證明**此前已流通的專有名詞用字（須提供**讀音**）；”

字符增收原則修訂 闡釋

- 要求更嚴
- 有關“此前已流通的專有名詞用字”的要求：

	舊版	新版
“此前已流通”？	須提供 依據	能(提供)證明
提供讀音？	無要求	須提供

謝謝！