

立法會資訊科技及廣播事務委員會

香港增補字符集

引言

本文件旨在解釋制定香港增補字符集的目的、字符集的收字原則及其他有關資料。

制定香港增補字符集的目的

2. ISO 10646 國際編碼標準以及香港常用的大五碼電腦中文編碼並未包納一些香港特區政府及市民在中文電子通訊中有需要使用的字符。由於缺乏一個共通的中文界面，出現了亂碼的情況（詳情請參考附件一）。制定香港增補字符集的目的，是收納這些香港電腦用戶有實際需要使用而暫未納入常用編碼標準內的中文字符，以提供一個共通的中文界面，方便大家能準確地以中文進行電子通訊。

附件一

3. 在一九九五年，香港政府為方便各部門交換電子檔案，在大五碼使用者造字區建立了一套名為政府通用字庫的中文字符集，收錄一些香港的特有用字，以補充大五碼基本字庫的不足。初期政府通用字庫只供政府內部使用，後來為方便市民下載政府文件，遂把政府通用字庫放在政府網站，供市民下載。

4. 一九九八年四月，資訊科技署與法定語文事務署合作，修訂政府通用字庫，並向各界收集中文電腦用字，編製一套共用的增補字符集，目的是提供一個共用的中文電腦界面，供政府部門及市民交換電子資料時使用。由於經增訂的字符集包羅從社會各界收集得來的字符，可作為社會人士共用的字符集，因而取名香港增補字符集。

香港增補字符集的制定過程及收字原則

5. 資訊科技署及法定語文事務署於一九九八年向政府部門及外間機構收集中文電腦用字後，先由法定語文事務署逐一審查並剔除重複的字符，然後翻查權威字典（包括《康熙字典》、《漢語大字典》、《漢語大詞典》和《中華字海》）、粵方言字典和文獻、ISO 10646 標準文件，以確定哪些字可以在字典找到、哪些字是約定俗成的粵方言用字、哪些字已包括在 ISO 10646 內。刪去來歷及用途不明的字符後，最後得出約一千七百個新字符。另外，該署也重新覆檢政府通用字庫，經整理後，將原有的 106 個字符刪除。新的香港增補字符集因此合計有 4 702 個字符。

6. 除此以外，資訊科技署亦成立了由學術界、語文及語言學學會、資訊科技業界及出版界的人士所組成的中文界面諮詢委員會(以下簡稱“中諮會”)，就向國際標準化組織提交本地中文電腦用字、在香港推廣

附件二

ISO 10646 國際編碼標準及協調香港各界人士有關本地中文電腦用字的需求向資訊科技署署長提供意見。中諮會的成員名單見附件二。中諮會及屬下兩個工作小組已召開多次會議，詳細討論增補字符集的收字原則、編碼方案和其他細節。視乎需要，中諮會的討論事項，會向資訊科技及廣播局轄下的資訊基建諮詢委員會匯報，以徵詢該委員會的意見。中諮會的討論文件亦放置在政府網頁上，供市民參考，並歡迎市民提出意見。

7. 香港增補字符集的收字原則，是每字須有來歷，並且必須是香港特區政府或市民有需要使用的字符。由於香港增補字符集將來會與 ISO 10646 接軌，因此審核收集得來的字符時，須參考 ISO 10646 的字形等同規則。香港增補字符集採取的收字原則，已經中諮會及轄下工作小組通過接納。

8. 經修訂的字符集及編碼於一九九九年九月中諮會會議上獲委員認可。字符集的規格文件於一九九九年九月二十八日正式公布，並於同日放置在政府網頁上供市民下載。香港增補字符集的規格公布後，軟件生產商可根據這套規格製作有關字型、輸入法、手寫板等產品。為了方便市民參考這套字符集，資訊科技署製作了一套字型和輸入法，放在政府網頁上供市民下載 (www.digital21.gov.hk/chi/hkscs)。

香港增補字符集收納粵方言用字的原因

9. 香港增補字符集收錄了一些粵方言用字。這些字主要由司法機構、警務處、律政司、語言學學會，以及大專院校提交，部分可在粵方言字典或學術著作中找到。字符集收納粵方言用字的原因之一，是方便司法機構記錄案件審訊過程，以及方便警務處和其他執法機關錄取口供之用。由於司法機構和執法機關都是採用逐字記錄的方式記錄審訊過程和口供，因此必須使用這些粵方言用字，才能提供完整的記錄。此外，粵方言研究是一門學科，不少語文學會會員和專上院校師生都以這個課題進行學術研究工作。若字符集不收錄粵方言用字，要刊印有關的論文著述，會有一定困難。把粵方言用字收納在字符集內，並非鼓勵使用這類字。字符集有如一個工具箱，放置着各種工具，不同使用者可因應各自的情況和需要，選取合適的工具使用。

香港增補字符集收納字典找不到的字的原因

10. 字符集中有一些字在字典找不到。這些字大部分來自入境事務處、公司註冊處、稅務局和地政總署的資料庫，並經證實用於現有的人名、公司名和地方名。由於人名、公司名和地方名用字會出現在各類證件、合約和法律文件上。這些文件出現人名、公司名和地方名時，必須使用該人或公司登記時使用的字，因此，為了應用上的需要，這些字也須納入字符集內。香港特區政

府和中諮會現正研究進一步改善現時收集、審核和管理中文電腦用字的機制，希望一方面能夠滿足市民和政府部門的電腦用字需求，另一方面可以避免增收不必要的字符。

香港增補字符集的管理機制

11. 中諮會已制定有關香港增補字符集的管理機制。根據該管理機制，中諮會及轄下兩個工作小組將負責以下工作：

- 建議審核和收納字符的準則；
- 審核香港增補字符集增收字符的申請；
- 為香港增補字符集增收字符編配內碼；
- 向市民發布香港增補字符集的修訂版本和公布有關消息；以及
- 就向國際標準化組織提交字符供納入 ISO 10646 國際編碼標準，和審核該等字符的事宜，提供意見。

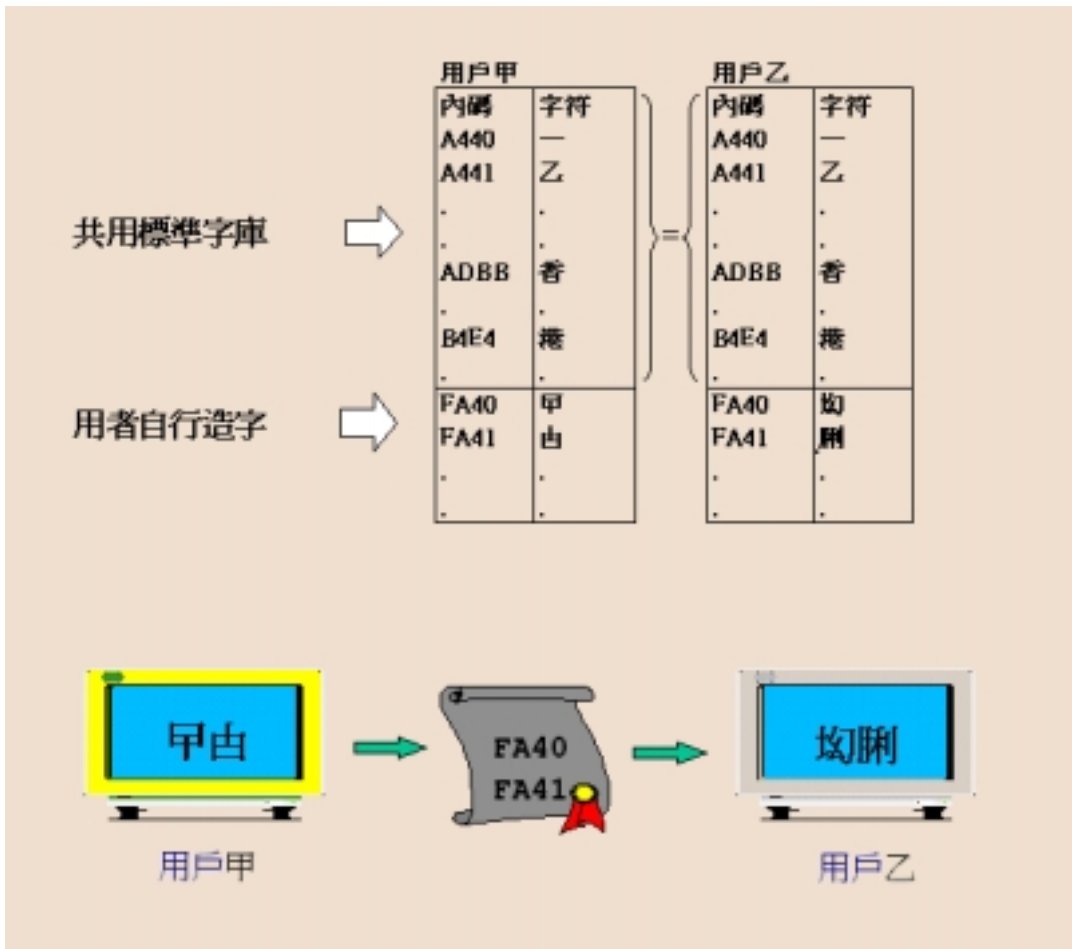
12. 香港增補字符集的管理機制綱要亦已在資訊科技及廣播局轄下的資訊基建諮詢委員會會議上討論，並獲得各委員支持。

13. 目前，中諮會轄下工作小組正制定管理機制的運作細則，預計可在二零零零年四月完成有關工作。

資訊科技署
二零零零年一月

缺乏共通中文界面所帶來的亂碼問題

要把資料存入電腦，必須根據某種預先設定的編碼方案把資料編碼。就中文資料而言，現時有多種編碼方案，包括大五碼、國標碼及 ISO 10646 國際編碼標準。上述各個編碼方案都未能把一些香港有需要使用的中文字符包括在內。這些字符之中，很多都是香港獨有的，包括人名、地名和粵方言用字。電腦用戶如需要在電腦上使用這些字符，唯有自行在使用者造字區為這些字符編配內碼。這對獨立運作的電腦而言，不會構成問題，但假如這些電腦互相通訊和交換資料，由用戶自行設定內碼的字符便會引起種種問題。在以下的例子，用戶甲和用戶乙在造字區用相同的碼位各自造了不同的字時，用戶甲的造字“甲由”傳送到用戶乙的電腦時，會錯誤地顯示為“坳廁”。



中文界面諮詢委員會委員名單

資訊科技署署長〔主席〕

巫偉明先生

李超倫先生

姜永正博士

姚德懷先生

陳真良博士

陸偉芃先生

陸勤博士

陸鏡光博士

張群顯博士

曾協泰先生

鄒嘉彥教授

鄭小康先生

鄭家安先生

潘應麟先生

戴澤棠先生

法定語文事務署代表

資訊科技及廣播局代表