

# 開放式中文字的組合

李果正 April 21, 2004

這個𪛗字是什麼意思。你的字典查得到嗎？你的電腦有辦法輸入這個字嗎？

以下的這些字呢？（看不清楚的話，請放大看仔細）

𪛗 𪛘 𪛙 𪛚 𪛛

我們需要這些字嗎？一般使用當然是不會需要，但學術研究會需要，一般文件有時也會需要，例如：

『這個𪛗字，是什麼意思，為什麼我的字典查不到？』

總不能限制文章不能這樣寫吧？

那麼使用 Unicode 不就得了？那麼，也請你告訴大家這些字的 Unicode 碼是什麼，在什麼字型上會有？

方塊字可以組合嗎？可以的，例如𪛗字，我們可以使用三個『用』字來組合，有這樣的字型技術嗎？有的，他稱為 stroke based font。但是目前表現出來的字型品質及視覺上的感覺會覺得生硬了些。目前已經有商業字型出現，也有人提議製定這種字型規格的標準，請參考：

<http://mail.nl.linux.org/linux-utf8/2003-04/msg00190.html>

商業字型請參考：

<http://www.bitstream.com/categories/developer/devfonts/sbfonts.html>

在字型技術上用心，例如 stroke based font，是一種方式，但從電腦更底層來設法似乎也是一種方式，也就是說在電腦系統裡頭不是直接的編碼對應，而是分成兩個層次的對應，第一個層次是 stroke 的對應，第二層次是 stroke

和組成字時的編碼的對應。爲什麼要這麼麻煩呢？因爲這樣才能和現有的編碼溝通，在第二層次的時候就可以和 Unicode 或其他現有的編碼來交換。

那麼總是會有目前現有編碼對應不到的字，就正如上面所提到的字，因此目前的編碼形式可能並不適合方塊字。

暫時的替代方案，可能就是把中央研究院文獻處理研究室的缺字方案 port 到 Linux 系統，要改的可能從 glibc 著手會比較好，但從應用軟體上著手也是一種方向，例如 OOo 及 L<sup>A</sup>T<sub>E</sub>X。L<sup>A</sup>T<sub>E</sub>X 的部份是沒有問題，只是規格、標準的制定而已，就如同這篇文章，輕易就可以把缺字嵌入，轉成 HTML 格式的話，則會自動轉換成 bitmap 圖檔。當然，這樣畢竟不是究竟的辦法，因爲其他的系統或軟體會無法製做。

中央研究院文獻處理研究室：

<http://www.sinica.edu.tw/~cdp/>

很可惜的，Richard S. Cook 在 2002 年的 Proposal，想把 CDP 的組合字的 operator 納入 Unicode，但並沒有成功：

<http://std.dkuug.dk/JTC1/SC2/WG2/docs/n2480.pdf>