

異體字、罕見字的處理

很久以前，看網路新聞，看到游錫「方方土」，王建「火宣」我很納悶，這是什麼東西。

在做通訊錄的時候，我們在處理人名，常常會遇到一些稀奇古怪的字，一種可能是我們的學養不夠，所以沒有認識這個字，如果這個字是正常的字，但是我們不會念，用手寫輸入法來處理就可以了。雖然我提過有線上的手寫輸入，但是它的字不夠多，往往手寫了，還是沒有相應的字出來，我還是比較喜歡搜狗中文輸入法附帶來的手寫輸入，寫進去，有就是有，沒有的話，就得從另一個角度來思考，另找辦法了。如果可以這樣找到字，我們就不用把「游錫堃」像過去必須寫成「游錫【方方土】」了

這些字有些是古字，有些是現代倉頡所造的字，用拼音（因為沒有看過這個字，不曉得讀法，有邊讀邊，沒邊讀中間當然找不到）、手寫都沒有辦法輸入。這些字，替文學家增添了许多意境。過去戶政事務是用手寫的，所以字隨便父母要什麼，就可以寫什麼。但是到了電腦時代，就沒有了這個字了。這些事，帶帶給了電腦輸入種種的難題。

好在我們的中央研究院在研究古書的時候，也碰到過這樣的問題，已經造出來許多這樣的字來，叫做「異體字」，行政院主計處電子處理資料中心也必須有套解的方案，才可以解決戶政上那些種種罕見的名字。

舉個例來說明，「瑩」這個字，有邊讀邊，沒變讀中間的原則，拼音「yin」找不到這個字；手寫也沒有這個字，就可以用下列的方式來尋找這個字。

<http://chardb.iis.sinica.edu.tw/>



在這個網路上去找。我選擇「部件」來找，意思是我不用「玉」字部首來找，這樣出來的字太多。看看有字裏有「瑩」的字有多少？



好極了，輸入「瑩」之後，找到這麼多字裏有「瑩」的，我所要的就是左邊第一個，可是這是一個圖畫，沒有辦法那裡做一般的文字處理。

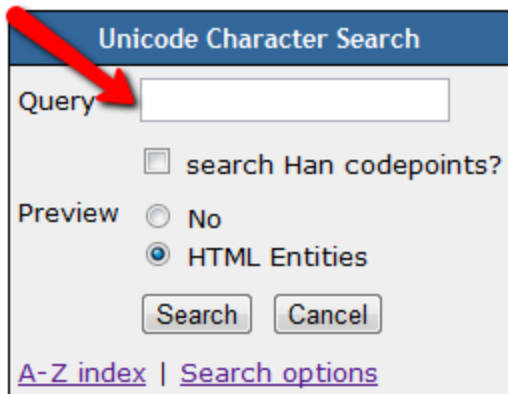
於是就點擊這個「瑩」字，到下一個網站，看到這個「字碼」



點擊這個字碼，就得到這個字的 unicode 碼 是「u+3f06」.

再到這個網站去找這個字的可編排的格式

<http://www.fileformat.info/info/unicode/char/search.htm>



Unicode Character Search

Query

☐ search Han codepoints?

Preview ☐ No
☒ HTML Entities

[A-Z index](#) | [Search options](#)

就在這個空格裏填入「u+3f06」, 點擊 search, 到達下一個網頁。



Unicode Character Search

Query:

☐ search Han codepoints?

Preview ☐ No
☒ HTML Entities

[A-Z index](#) | [Search options](#)

Results

Search: u+3f06
1 hits

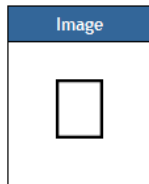
Are you looking for Unicode character U+3F06: U+3F06?

Matching Text	Character	Description	Preview
U+3F06	U+3F06	U+3F06	璚

這時候, 這個「璚」字就在箭頭 1 的地方了, copy and paste 到其他檔案就可以了。當然先決條件是個人的電腦, 必須有這個 unicode 的字型才可以。

如果還不知足, 就點擊箭頭 2 所在的「U+3F06」到下一個網頁

Unicode Han Character 'U+3F06' (U+3F06)



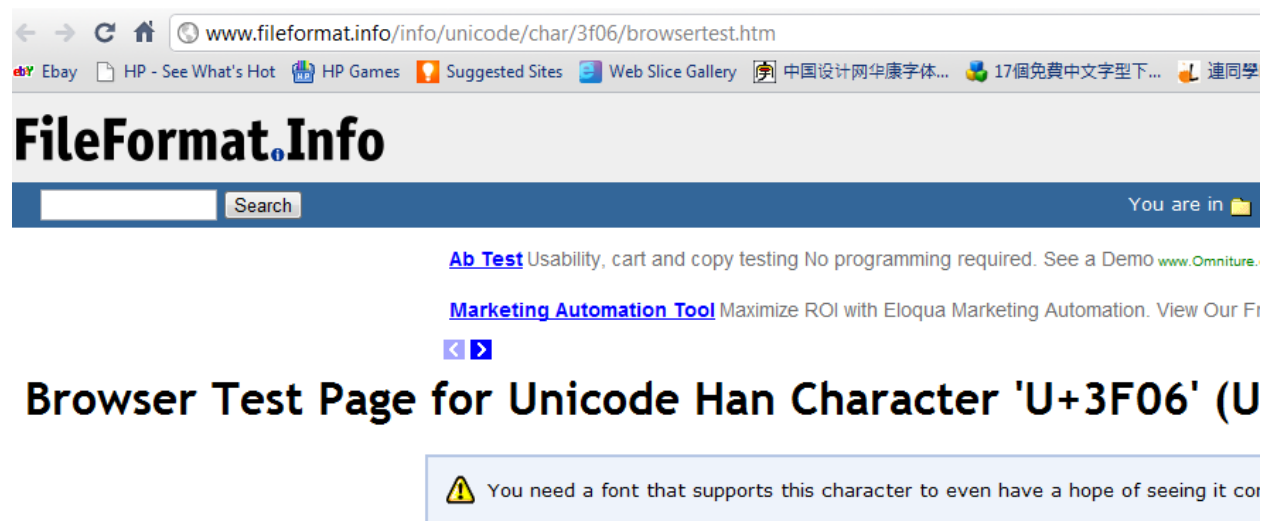
[Browser Test Page](#) | [Outline \(as SVG file\)](#) | [Fonts that support U+3F06](#)

Unicode Han Data	
kRSUnicode	96.15
kHKSCS	9463

然後再點擊箭頭所指「Browser Test Page」

就會帶到下列網址，得到更多的資料，copy and paste 箭頭所指的字就可以運用了。

<http://www.fileformat.info/info/unicode/char/3f06/browsertest.htm>



The screenshot shows the FileFormat.Info website. The browser address bar displays the URL: www.fileformat.info/info/unicode/char/3f06/browsertest.htm. The page title is "Browser Test Page for Unicode Han Character 'U+3F06' (U+3F06)". A warning message states: "You need a font that supports this character to even have a hope of seeing it cor". Below the warning, there is a table showing the character's appearance in different encodings. A red arrow points to the "Input Test" section, which shows the character "璚" in an input field.

Should look like	UTF-8	Hex escape	Decimal escape
	璚	璚	璚

See the [local font list](#) tool and [font support page](#) to see if you have a font that supports it!

Should look like	UTF-8	Hex escape	Decimal escape
	璚	璚	璚

Input Test	
As UTF-8 in an input field	璚

另一個字「滌」，可以用同樣的辦法來取得。

這個也是一個很不錯的網站，可以處理這個異體字，是行政院主計處電子處理資料中心的版權。

<http://www.cns11643.gov.tw/AIDB/welcome.do>

<http://www.cns11643.gov.tw/cgi-bin/ttf2png?page=13&number=2c2d&face=sung&fontsize=26>

<http://www.ceec.edu.tw/Download/Download.htm>

下面這個，可能會很有用，但是我沒有中文視窗的界面，所以沒有辦法來說到底可不可以用。

<http://www.ceec.edu.tw/Download/Download.htm>

【下載造字】適用於有安裝『全字庫軟體包』的作業系統(如 Windows 2000、XP、2003) Windows 2000、XP、2003內預設沒有該字型 若您要下載此造字，請依下列步驟操作： 1、請點選這裡 2、於所出現的下載視窗中，選擇「開啟舊檔」 3、並於出現的「造字轉入及列印」的視窗中，點選下方的「加入」按鈕 4、電腦重新登入或重開機後，可用舊注音、舊倉頡、Big5或 Unicode 內碼輸入該造字 ◎下載之前請先確認是否已安裝[全字庫軟體包]	
【下載造字】適用於有安裝『全字庫應用工具4.0版』的作業系統(如 Windows 95、98、ME) 1、下載明體點矩陣字形、注音及倉頡輸入法，存放於 C:\CMEX 目錄 2、下載楷體 TrueType 字形，存放於 C:\CMEX 目錄(若不需楷體字形，可免下載) 3、執行「字形轉入工具」，以將字形轉入至電腦系統中 4、之後即可用舊注音、舊倉頡或 Big5內碼輸入該造字 ◎下載之前請先確認是否已安裝[全字庫應用工具4.0版]	

𪗇 (U+E4FB in EUDC)

這個字沒有辦法用上述的方法取得，可能不是unicode有的字，是自己創造出來的。就只好到台灣的大學入學考試中心那裡，找到了其造字檔。在那個檔裏面，找到了「𪗇」這個字。但是沒有可以用輸入法來得到這個字，只好用 character map來找到，copy and paste到自己所要的檔案去。這個方法，只有在自己的電腦上有用，檔傳給了別人，可能就沒有 這個字形在其電腦裏，可能會顯出一個點的記號而已。

[http://www.ceec.edu.tw/Download/Word/EUDC\(1000201\).zip](http://www.ceec.edu.tw/Download/Word/EUDC(1000201).zip)