

《神農本草經》舊版造字轉碼說明

中研院資訊所文獻處理實驗室

中研院史語所漢籍電子文獻工作小組

2008/3/25 陳建安 製作

一、《神農本草經》一書(神農本草經.xml)使用舊版造字 4 個，字頻 8 次，詳如附件一。這 4 個造字中，2 個可轉成 Windows XP 能顯示的字，字頻 5 次；另外 2 個字必須轉成構字式，字頻 3 次。

二、附件一的造字分析表說明如下：

甲、編號：Big5 造字空間為 6217 個，編號由 1 到 6217。

乙、造字：舊版造字

丙、字頻(txt)：造字在「.txt」文件的出現次數

丁、字頻(xml)：造字在「.xml」文件的出現次數

戊、Big5：造字的 Big5 碼

己、Unicode：造字所對應的 Unicode 碼

庚、WinXP：造字在 Windows XP 的對應字形

辛、構字式：Windows XP 若無對應字形，則改採用構字式

壬、備註凡例：

- 1、異體字問題：新版漢籍考量到使用者檢索及使用時的便利性，將用字原則改為除專詞等特殊情形之外，一律改用標準字呈現。如編號 4373 的「鼠△蟲」係「鼯」字之異體，故以「鼯」字取代。

三、Unicode 目前收錄的漢字總數為 70194，分屬於三個不同區段，詳如表一。目前 Windows XP 只支援 CJK 認同表意文字區的 20902 個字，內碼為 4E00-9FFF。所以造字編號 5335 的「𧈧」字，Unicode 編碼為 459F，由於 Windows XP 並不支援，仍須使用構字式「亡𧈧」。

表一、Unicode 的字數及編碼區段

Unicode	字集子集合	新增字數	新增編碼區段	總字數	WinXP
1.1 版	CJK 認同表意文字區	20902	4E00-9FFF	20902	支援
3.0 版	CJK 認同表意文字擴充 A 區	6582	3400-4DFF	27484	不支援
3.1 版	CJK 認同表意文字擴充 B 區	42710	20000-2A6D6	70194	不支援

附件一、《神農本草經》造字分析表

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
3787	豕	1	1	8152	51A2	豕		
3796	決	4	4	815B	51B3	決		決，異體字問題。
4373		2	2	84E7			鼠△𧈧	𧈧，異體字問題。
5335	𧈧	1	1	8AFB	459F		亡△虫	