

《黃帝針灸甲乙經》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、《黃帝針灸甲乙經》一書(黃帝針灸甲乙經.xml)使用舊版造字 11 個，字頻 16 次，詳如附件一。這 11 個造字中，8 個可轉成 Windows XP 能顯示的字，字頻 12 次；另外 3 個字必須轉成構字式，字頻 4 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 3824 的「叙」係「敘」字之異體，故以「敘」字取代。又編號 2717 的「人叢米」字，係「糴」字之簡體字，亦以正體的「糴」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 5318 的「蔭」字，Unicode 編碼為 4516，由於 Windows XP 並不支援，仍須使用構字式「𠂇陵」。

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、《黃帝針灸甲乙經》造字分析表

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
432	業	1	1	FCD7			業	
844	膠	2	2	8E7A	9ACE	膠		
1653	腠	1	1	93B4	8142	腠		
2325	睨	2	2	97E0	77A4	睨		
2717		2	2	9A6F			人𠂇米	糴，異體字問題。
3824	叙	1	1	8177	53D9	叙		敘，異體字問題。
3967	炳	3	3	8269	712B	炳		
4768	妬	1	1	8779	59AC	妬		妒，異體字問題。
5302	昏	1	1	8ADA	662C	昏		
5318	蔭	1	1	8AEA	4516		艹𠂇陵	
5724	穡	1	1	8DA9	7A38	穡		