

《衛生家寶產科備要》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

2008/2/27 陳建安 製作

一、《衛生家寶產科備要》一書(衛生家寶產科備要.xml)使用舊版造字 25 個，字頻 44 次，詳如附件一。這 25 個造字中，19 個可轉成 Windows XP 能顯示的字，字頻 35 次；另外 6 個字必須轉成構字式，字頻 9 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、校對問題，新舊版缺字並存：漢籍在舊轉新的過程中，以人工修改方式取代舊漢籍造字，雖然修改的結果是正確的，但是只修改到部分造字，沒有全部修改到，導致舊漢籍造字在新版漢籍檔案中，一部分已被取代成 Unicode 或構字式，另一部分仍為舊漢籍造字的並存情形。處理方式為配合漢籍工作室的修改，將沒被修改到的舊漢籍造字，以對應的 Unicode 或構字式取代。如編號 4750，舊漢籍造字為「𠂔」，比對 txt 檔及 xml 檔後，發現原先在 txt 檔使用舊漢籍造字「𠂔」者，在 xml 檔中部分已人工修改為「𠂔」，故配合修改，將編號 4750 的「𠂔」字，手動取代為「𠂔」。

2、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 3789 的「冲」係「沖」字之異體，故以「沖」字取代。又如編號 3797 的「凭」係「憑」字之簡體，故以「憑」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 1310 的「𧈧」字，Unicode 編碼為 45EB，由於 Windows XP 並不支援，仍須使用構字式「虫𧈧麻」。

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

四、 txt 檔與 xml 檔頻次不同且造字分析表「備註」欄有備註問題情況者，詳細取代內容可參考附件二的手動取代表。附件二的手動取代表的欄位說明如下：

- 甲、編號：Big5 造字空間為 6217 個，編號由 1 到 6217。
- 乙、取代：實際使用字形與舊漢籍造字不同時，於此欄列出。
- 丙、檔案原文摘錄：摘錄檔案(衛生家寶產科備要.xml)中包含該造字的文句段落，其中紅字底線的部分，為需要手動取代的原文。
- 丁、手動取代結果：變更過後的檔案內容，其中藍字底線的部分，為修改後的結果。

附件一、《衛生家寶產科備要》造字分析表

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
1310	𧈧	2	2	9175	45EB		虫𧈧麻	
1890	𧈧	1	1	9545	44D7		𧈧𧈧𧈧	
2505	𧈧	3	3	98F7	98DC	𧈧		
3789	冲	1	1	8154	51B2	冲		冲，異體字問題。
3797	凭	1	1	815C	51ED	凭		憑，異體字問題。

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
3924	朥	1	1	81FD	671E	朥		
3931	椀	11	11	8245	6900	椀		
3968	煨	1	1	826A	7145	煨		
4007	癰	1	1	82B3	7667	癰		
4014	蓋	3	3	82BA	8462	蓋		蓋，異體字問題。
4063	藜	1	1	82EB	7CF5	藜		
4069	綑	1	1	82F1	7DB3	綑		
4080	羣	1	1	82FC	7FA3	羣		
4115	藁	1	1	8360	85C1	藁		
4195	鉢	2	2	83D2	9262	鉢		
4507	吳	1	1	85D0	5434	吳		吳，異體字問題。
4525	𠵽	1	1	85E2	54EF	𠵽		
4750	𠵽	10	1	8767			𠵽	𠵽，校對問題，新舊版缺字並存。
4799	顙	1	1	87BA	9856	顙		
4803	腭	1	1	87BE	816D	腭		
5034	弃	2	2	8949	5F03	弃		棄，異體字問題。
5149	胥	1	1	89DE	2676B		要△月	腰，異體字問題。
5218		3	3	8A64			𠵽→覓、𠵽	寬，異體字問題。

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
5335	𧈧	1	1	8AFB	459F		亡𧈧虫	
5398	瀘	1	1	8B7B	6E7B	瀘		

附件二、《衛生家寶產科備要》手動取代表

編號	取代	檔案原文摘錄	手動取代結果
4750	𠄎	蓬莪 <u>𠄎</u> 𧈧戍	蓬莪 <u>𠄎</u> 𧈧戍