

《子和醫集》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
2008/3/13 黃榮順 製作

一、《子和醫集》一書(子和醫集.xml)使用舊版造字 34 個，字頻 59 次，詳如附件一。這 34 個造字中，26 個可轉成 Windows XP 可顯示的字，字頻 45 次；另外 8 個字必須轉成構字式，字頻 14 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、重碼問題，採用部件外字，不轉碼：舊漢籍造字與構字式部件外字，皆是建立在 Big5 的使用者造字區中，兩者所使用的碼區是重疊的，在同時使用構字式和舊漢籍造字的新版漢籍中，這些重疊的字碼可能被用以表達兩個不同字形，我們稱之為「重碼問題」。然而漢籍在舊轉新的過程中，若已將原舊漢籍造字的部分全數改為 Unicode 或構字式，則在 xml 出現的造字頻次皆採用部件外字，此時便不需轉碼。如編號 4817，字碼 87CC，對應的舊漢籍造字為「寶」，部件外字為「ㄨ」，比對原書的結果，xml 檔的頻次皆採用部件外字「ㄨ」，故不需轉碼。
- 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 2936 的「竒」係「奇」字之異體，故以「奇」字取代。
- 3、Unicode 字型呈現差異：Unicode 字型與舊漢籍造字有些微差異，但只是字體風格差異，實際上仍為同一個字，因此仍取 Unicode 字型。如編號 5131 的「窻」字，Unicode

字型呈現為「窻」，實際上仍為同一字。

4、待造字：Unicode 及漢字構形資料庫皆未收錄的舊漢籍造字，正在等待補造字中，所以「造字」欄空白無法看到字形。如編號 2962 的「𠄎𠄎𠄎黃」。

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。所以如造字編號 2276 的「𦍋」字，Unicode 編碼為 26369，由於 Windows XP 並不支援，仍須改用構字式「羊𠄎古」。

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、「子和醫集.xml」發現之未確定字形「●」共 1 處，修正情形詳列附件二，附件二「子和醫集.xml」「●」修改表體例說明如下：

甲、檔案原文摘錄：摘錄檔案(子和醫集.xml)中包含「●」的文句段落，其中紅字底線的部分，為需要手動取代的原文。

乙、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。

五、「子和醫集.xml」缺字外發現之錯字共 4 處，修正情形詳列附件三，附件三「子和醫集.xml」缺字外錯字修改表體例說明如下：

甲、檔案原文摘錄：摘錄檔案(子和醫集.xml)中包含該錯字的文句段落，其中紅字底線的部分，為需要手動取代的原文。

乙、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、「子和醫集.xml」造字分析表

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
1591	羗	1	1	9354	7F97	羗		
1784	萼	1	1	9478	855A	萼		
2276	羗	1	1	97AF	26369		羊厶古	
2434		2	2	98B0			卩厶道	隨，異體字問題。
2931		1	1	9BCA			此厶白	皆，異體字問題。
2936	奇	1	1	9BCF	7AD2	奇		奇，異體字問題。
2962		1	1	9BE9			艹厶黃	待造字。
3796	決	8	8	815B	51B3	決		決，異體字問題。
3850	姪	1	1	81B3	59D9	姪		
3876	廻	1	1	81CD	5EFB	廻		廻，異體字問題。
3951	湏	3	3	8259	6E4F	湏		
3996	疎	2	2	82A8	758E	疎		
4036	窰	3	3	82D0	7AB0	窰		
4045	箒	1	1	82D9	7B92	箒		
4085	耻	1	1	8342	803B	耻		恥，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4091	臯	1	1	8348	81EF	臯		Unicode 字型差異。
4172	麒	1	1	83BB	8E77	麒		
4184	迹	2	2	83C7	8FF9	迹		跡，異體字問題。
4220	穎	2	2	83EB	9834	穎		穎，異體字問題。
4262		1	1	8456			麥△丐	麵，異體字問題。
4278	亘	1	1	8466	4E98	亘		
4716	埜	3	3	8745	57FF	埜		
4767	懷	4	4	8778	61B9	懷		
4803	腭	1	1	87BE	816D	腭		
4812	羗	3	3	87C7	26351		羗△、	
4817	寶	16	1	87CC	5BF3	寶		ㄣ，重碼問題，採用部件外字，不轉碼。
4848	輩	1	1	87EB	8EF0	輩		輩，異體字問題。
5027	牕	1	1	8942	724E	牕		
5131	窻	1	1	89CC	7ABB	窻		Unicode 字型差異。
5386	塼	3	3	8B6F	365B		土△專	
5439	栢	1	1	8BC6	6822	栢		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
5482	璽	1	1	8BF1	8812	璽		
5533		2	2	8C65			穀△米	穀，異體字問題。
5590	朶	1	1	8CC0	6736	朶		朶，異體字問題。

附件二、「子和醫集.xml」 「●」修改表

檔案原文摘錄	手動取代結果
夏(j●á 夾)	夏(jiá 夾)

附件三、「子和醫集.xml」缺字外錯字修改表

檔案原文摘錄	手動取代結果
驚、?? 殺肉	驚、 殺肉
原作「  𠃉 一 𠃉 𠃉」	原作「笱」
石盤△礪	石△盤礪
缺竦而石盤△礪	缺竦而石△盤礪