

《名醫類案》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
2008/3/11 陳建安 製作

一、《名醫類案》一書(名醫類案.xml)使用舊版造字 75 個，字頻 147 次，詳如附件一。這 75 個造字中，57 個可轉成 Windows XP 能顯示的字，字頻 106 次；另外 18 個字必須轉成構字式，字頻 41 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、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 1332 的「扌𠂔𠂔」係「攜」字之異體，故以「攜」字取代。

2、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 4158 的「譌」字，Unicode 字型呈現為「譌」，實際上仍為同一字。

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 790 的「𠂔」字，Unicode 編碼為 3A49，由於 Windows XP 並不支援，仍須使用構字式「目𠂔夾」。

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、對於在缺字之外發現的錯字，其詳細內容可參考附件二的校出錯字表。附件二《名醫類案》校出錯字表的說明如下：

- 甲、檔案原文摘錄：摘錄檔案(名醫類案.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、《名醫類案》造字分析表

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
790	映	1	1	8E44	4039		目 𠃉 夾	
842	𠃉	1	1	8E78	3AB7		𠃉	
1077	𠃉	1	1	8FE8	7A3E	𠃉		
1151	𠃉	3	3	9073	5273	𠃉		
1332		1	1	91AD			扌 𠃉 𠃉	攜，異體字問題。
1394	𠃉	3	3	91EB	35D6		口 𠃉 𠃉	啗，異體字問題。
1616	𠃉	1	1	936D	284E6		辶 𠃉 里	
1664	腸	1	1	93BF	8193	腸		腸，異體字問題。
2833	刃	2	2	9B46	5204	刃		刃，異體字問題。
3004	籊	1	1	9C54	7BD0	籊		
3176		1	1	9D63			纟 𠃉 段	緞，異體字問題。
3799	刮	1	1	815E	5226	刮		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
3809	匱	1	1	8168	5333	匱		
3836	喻	1	1	81A5	564F	喻		
3843	堦	1	1	81AC	5826	堦		
3846	堦	1	1	81AF	58FB	堦		Unicode 呈現差異。
3858	寃	1	1	81BB	5BC3	寃		
3865	瀟	1	1	81C2	24162		灬 ㄩ 蕭	
3876	廻	3	3	81CD	5EFB	廻		迴，異體字問題。
3877	迺	3	3	81CE	5EFC	迺		迺，異體字問題。
3910	擘	1	1	81EF	64E5	擘		
3927	柏	1	1	8241	6855	柏		
3976	犁	1	1	8272	7282	犁		
3977	狗	1	1	8273	72E5	狗		
4013	盃	3	3	82B9	76CC	盃		
4032	秣	1	1	82CC	2578A		禾 ㄩ 未	秣，異體字問題。
4041	豎	1	1	82D5	7AEA	豎		
4093	舐	1	1	834A	445B		舌 ㄩ 氏	舐，異體字問題。
4138	袞	1	1	8377	88A0	袞		
4140	袞	2	2	8379	88B5	袞		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4158	譌	1	1	83AD	8B4C	譌		Unicode 呈現差異。
4184	迹	2	2	83C7	8FF9	迹		
4186	遡	2	2	83C9	9061	遡		
4190	邳	1	1	83CD	90C4	邳		
4218	菲	1	1	83E9	97EE	菲		
4221	頰	2	2	83EC	9833	頰		
4226	凜	2	2	83F1	98E1	凜		
4251	鰕	2	2	844B	9C15	鰕		
4256	鷄	1	1	8450	9DC4	鷄		
4274		11	11	8462			扌厶友	拔，異體字問題。
4275	义	1	1	8463	4E49	义		又，異體字問題。
4283	戲	1	1	846B	622F	戲		戲，異體字問題。
4299	焯	2	2	847B	242F2		火厶芮	
4325	賒	1	1	84B7	8CD6	賒		
4423	醫	3	3	855A	81CB	醫		
4437	廩	2	2	8568	5EEA	廩		
4442	潛	1	1	856D	6F98	潛		Unicode 呈現差異。
4450	翔	1	1	8575	524F	翔		翔，異體字問題。

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4482	厯	1	1	85B7	53AA	厯		
4496	段	5	5	85C5	53DA	段		段，異體字問題。
4504	𠵱	2	2	85CD	541A	𠵱		
4525	𠵱	1	1	85E2	54EF	𠵱		
4584	喜	1	1	865E	569E	喜		
4746		4	4	8763	28932		𠵱牙女厯	釐，異體字問題。
4803	𠵱	2	2	87BE	816D	𠵱		
4902	充	2	2	8862	5156	充		
4979	錫	1	1	88D1	296FF		食 𠵱 易	
5008	𠵱	1	1	88EE	5E5E	𠵱		
5045	𠵱	1	1	8954	9901	𠵱		
5046	𠵱	2	2	8955	5F5C	𠵱		
5072	𠵱	1	1	896F	6244	𠵱		
5131	窻	1	1	89CC	7ABB	窻		Unicode 呈現差異。
5219	妃	4	4	8A65			女 𠵱 巳	妃，異體字問題。
5315	𠵱	4	4	8AE7	3CC2		𠵱 𠵱 公	
5335	𠵱	2	2	8AFB	459F		亡 𠵱 虫	
5387	𠵱	1	1	8B70	20374		𠵱 𠵱 夙 心	

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
5406	尻	1	1	8BA5	51E5	尻		
5408	絛	4	4	8BA7	7D4D	絛		
5579	窰	2	2	8CB5	5D08	窰		
5595	塚	1	1	8CC5	579C	塚		
5666	汚	19	19	8D4D	6C5A	汚		汚，異體字問題。
5693	𪗇	1	1	8D68	9E7B	𪗇		
5694		1	1	8D69			𠂇𠂇𠂇𠂇	概，異體字問題。
5703	𪗈	1	1	8D72	9F53	𪗈		
5705	廐	2	2	8D74	5ED0	廐		廐，異體字問題。

附件二、《名醫類案》校出錯字表

檔案原文摘錄	手動取代結果
茲??第以飲食勞倦	茲第以飲食勞倦