

《名醫別錄》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
2008/3/14 黃榮順 製作

一、《名醫別錄》一書(名醫別錄.xml)使用舊版造字 60 個，字頻 113 次，詳如附件一。這 60 個造字中，38 個可轉成 Windows XP 可顯示的字，字頻 78 次；另外 22 個字必須轉成構字式，字頻 35 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，舊版漢籍錯字，可用程式全部取代：在舊版漢籍電子文獻中即存在的錯字，因校對時的疏漏而未更正，持續留存在新版漢籍電子文獻中；若該造字的所有頻次，皆屬於錯誤使用的錯字情形，可以用程式全部取代為正確字形。如編號 2461 的「𠂔𠂔𠂔忽」字，原字應為「葱」。
- 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 4768 的「妬」係「妒」字之異體，故以「妒」字取代。又如編號 3954 的「𠂔𠂔𠂔𠂔」係「戢」字之簡體，故以「戢」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 4259 的「𠂔𠂔」字，Unicode 字型呈現為「𠂔𠂔」，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 2244 的「ノ 𧈧 虫」。

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 888 的「𣎵」字，Unicode 編碼為 3B88，由於 Windows XP 並不支援，仍須改用構字式「木 𣎵 奈」。

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、「名醫別錄.xml」造字分析表

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
639	𠂔	1	1	FE4A			𠂔	
690	𠂔	1	1	FE7D			𠂔	
888	𣎵	1	1	8EC8	3B88		木 𣎵 奈	
981	𣎵	1	1	8F66	72B2	𣎵		
1067	𣎵	1	1	8FDE	68CA	𣎵		
1650	𣎵	1	1	93B1	4417		月 𣎵 者	
1668	𣎵	1	1	93C3	81C8	𣎵		
1780	𣎵	1	1	9474	67E0	𣎵		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
1981	讐	2	2	95C2	8B90	讐		
1999	贊	1	1	95D4	8CDB	贊		
2044	踈	1	1	9642	8E08	踈		
2244		1	1	976D			ノ厶虫	待造字。
2260	癩	2	2	977D	3FD7		疒厶頰	
2386	瑯	5	5	985E	7458	瑯		
2461	葱	1	1	98CB	44E4		艹厶忽	葱，校對問題，舊版漢籍錯字，可用程式全部取代。
2496	鰈	1	1	98EE	4C5C		魚厶昔	
2539	胙	1	1	995A	43F6		月厶𠂔	
2628	莢	3	3	99D5	845C	莢		
2775	茱	1	1	9ACB	3B49		茱	
2863		2	2	9B64			艹厶友	茭，異體字問題。
3787	豕	1	1	8152	51A2	豕		
3805	勅	1	1	8164	52C5	勅		
3858	寃	1	1	81BB	5BC3	寃		
3954		2	2	825C			晋厶戈	戩，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
3966	烟	2	2	8268	70DF	烟		
3979	猫	5	5	8275	732B	猫		貓，異體字問題。
3996	疎	2	2	82A8	758E	疎		
4007	癰	6	6	82B3	7667	癰		
4017	蟬	1	1	82BD	880F	蟬		
4042	笋	4	4	82D6	7B0B	笋		
4063	蘖	2	2	82EB	7CF5	蘖		
4087	鰓	2	2	8344	4C47		魚𩺰旦	
4103	茝	1	1	8354	831D	茝		
4105	菓	1	1	8356	83D3	菓		
4113	藺	2	2	835E			艹厶間	
4190	邴	1	1	83CD	90C4	邴		
4193	鈎	8	8	83D0	920E	鈎		
4218	萑	1	1	83E9	97EE	萑		
4220	穎	2	2	83EB	9834	穎		穎，異體字問題。
4247	鮎	1	1	8447	9B8E	鮎		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4259	麿	3	3	8453	9E95	麿		Unicode 字型差異。
4262		1	1	8456			麥△𪛗	麵，異體字問題。
4267	鼈	4	4	845B	9F08	鼈		
4345		2	2	84CB			口△𪛗	待造字。
4437	廩	3	3	8568	5EEA	廩		
4485	廙	1	1	85BA	3553		厂△墨	
4768	妬	1	1	8779	59AC	妬		妒，異體字問題。
4789	蝮	1	1	87B0	878B	蝮		
4790	賺	1	1	87B1	81C1	賺		
4799	顚	1	1	87BA	9856	顚		
4910		5	5	886A			汙△𪛗	淫，異體字問題。
4930	岑	1	1	887E	5CBA	岑		
5104	礼	2	2	89B1	793C	礼		
5455	𪛗	1	1	8BD6	48DB		𪛗△𪛗	
5501	𪛗	1	1	8C45	4AA5		𪛗	
5652	𪛗	2	2	8CFE	27312		或△𪛗	

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
5697	𩇛	3	3	8D6C	7151	𩇛		煮，異體字問題。
5706	𩇜	1	1	8D75	654D	𩇜		敘，異體字問題。
5709	𩇝	1	1	8D78	872F	𩇝		
5783	𩇞	3	3	8DE4	3A8D		𩇞𩇞𩇞	