

《本草綱目》舊版造字轉碼說明

中研院資訊所文獻處理實驗室

2006/12/13 趙苑曲

一、《本草綱目》一書(本草綱目.xml)使用舊版造字 171 個，字頻 358 次，詳如附件一。這 171 個造字中，105 個可轉成 Windows XP 可顯示的字，字頻 246 次；另外 66 個字必須轉成構字式，字頻 112 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：造字在文件中的出現次數

丁、Big5：造字的 Big5 碼

戊、Unicode：造字所對應的 Unicode 碼

己、WinXP：造字在 Windows XP 的對應字形

庚、構字式：Windows XP 若無對應字形，則改採用構字式。

辛、備註凡例：

- 1、取代：用另一個字取代原造字，如編號 2434 的「𠂔𠂔」字，改用「隨」字取代。
- 2、錯字：用正確的字取代原造字，如編號 1660 的「𦵏」字，原字應為「𦵏」。
- 3、字形未定義，手動取代：未定義的字形，必須一一比對原書字形再以正確字形分別取代。如編號 309 的「𠂔」，屬於未定義的字形，比對原書後，以「𠂔」字取代。
- 4、構字式錯誤，手動取代：檔案中使用錯誤的構字式表達字形，改以正確的字形取代。如編號 781 的「𠂔」，在檔案中以「𦵏𠂔」表達「𦵏」字，可直接以「𦵏」字取代構字式「𦵏𠂔」。
- 5、重碼造字，手動取代：同時使用舊漢籍造字及構字式，導致一個字碼代表兩種字形的重碼問題，必須一一比對原書字形，以判斷應取何者之字形，取舊漢籍造字字形者再以手動取代。如編號 5351，舊漢籍造字字形為「𦵏」，構字式部件為「𦵏」，一一比對原書字形為「𦵏」者再以手動取代。
- 6、重碼部件，不轉碼：同時使用舊漢籍造字及構字式，導致一個字碼代表兩種字形的重碼問題，必須一一比對原書字形，以判斷應取何者之字形，取構字式部件者不轉

碼。如編號 5033，舊漢籍造字字形為「𦵏」，構字式部件為「𦵏」，比對原書發現書中皆取構字式部件字形「𦵏」，故不需轉碼。

7、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 4793 的「𦵏」字，Unicode 楷體為「𦵏」，明體為「𦵏」，實際上與「𦵏」字相同。

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 5335 的「𦵏」字，Unicode 編碼為 459F，由於 Windows XP 並不支援，仍須改用構字式「亡𧈧虫」。

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、造字分析表「備註」欄標示「手動取代」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：舊版造字字形

丙、部件：構字式字形，通常當部件使用。

丁、檔案原文摘錄：摘錄檔案(本草綱目.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。

戊、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、《本草綱目》造字分析表

編號	造字	字頻	Big5	Unicode	WinXP	構字式	備註
20	藪	1	FA53			竹 叢 救	
28	沢	1	FA5B	23CB5		氵 厶 只	
309	𠂇	2	FBF9			𠂇	「𠂇」，字形未定義，手動取代。
379	支	1	FCA2			支	
432	業	1	FCD7			業	
479	畐	1	FD47			畐	
488	毘	1	FD50			毘	
489	隶	1	FD51			隶	
632	炆	3	FE43			炆	
711	𠂇	1	FEB4			𠂇	
781	𠂇	1	FEFA			𠂇	「𠂇」，構字式錯誤，手動取代。
814	墓	2	8E5C	87C7	墓		
841	礪	1	8E77	7932	礪		
845	韶	1	8E7B			鼠 厶 召	
887	鷲	2	8EC7	9E0E	鷲		
893	戕	1	8ECD	5CE9	戕		
953	漫	2	8F4A	23E70		氵 厶 曼	
959	揄	1	8F50	3A49		扌 厶 俞	
977	麇	1	8F62			鹿 厶 君	
983	瘕	3	8F68	24DCD		疒 厶 其	
1008	玦	1	8FA3	739E	玦		
1056	阤	3	8FD3	9641	阤		
1067	綦	1	8FDE	68CA	綦		
1082	樛	2	8FED	3BC9		木 厶 庖	
1084	鷁	1	8FEF	4CDC		余 厶 鳥	
1146	蕘	5	906E	7506	蕘		
1151	剗	1	9073	5273	剗		
1285	礪	1	915C	40CB		石 厶 俞	
1310	蠲	2	9175	45EB		虫 厶 麻	
1341	豨	1	91B6	8C7C	豨		

編號	造字	字頻	Big5	Unicode	WinXP	構字式	備註
1365	櫛	3	91CE	6AC4	櫛		
1387	腎	1	91E4	7796	腎		
1429	荳	2	924F	4532		艹 厶 登	
1530	紕	5	92D6	7CFA	紕		
1557	羴	3	92F1			羊 厶 歷	
1643	蠶	1	93AA	27525		虫 厶 遺	
1660	賺	1	93BB	8181	賺		「賺」，錯字。
1736	蒺	1	9448	8453	蒺		
1755	蔘	2	945B			艹 厶 修	
1780	柠	1	9474	67E0	柠		
1882	蛞	1	94FC	86D2	蛞		
1976	蝨	1	95BD	45D0		施 厶 虫	
2044	踈	2	9642	8E08	踈		
2094	厯	1	9674	4365		𠂇 厶 歷	
2147	蟥	3	96CB	87EE	蟥		
2182	驪	1	96EE	4BBE		馬 厶 龍	
2188	離	1	96F4	4366		𠂇 厶 離	
2206	袄	1	9747	8884	袄		
2246	炆	1	976F	7079	炆		
2276	𦍋	1	97AF	26369		羊 厶 古	
2322	𦍋	1	97DD	4948		金 厶 莽	
2367	篁	1	984B	25C2D		竹 厶 差	
2369	舩	2	984D	26A34		舟 厶 同	
2374	冢	1	9852	51A1	冢		
2411	𡗗	1	9877	3C26		去 厶 欠	
2434	随	2	98B0			卩 厶 迨	「隨」，取代。
2449	雒	1	98BF			名 厶 隹	
2454	鍬	2	98C4	28AFC		金 厶 敕	
2539	胾	5	995A	43F6		月 厶 望	
2610	鰓	2	99C3	9C53	鰓		
2643	獫	1	99E4	2486E		犛 厶 然	
2779	莖	2	9ACF	84C8	莖		

編號	造字	字頻	Big5	Unicode	WinXP	構字式	備註
2863	芟	2	9B64			艹 ㄉㄨㄞ 友	「芟」，取代。
2916	檄	1	9BBB	6A27	檄		「檣」，錯字。
3006	蠟	1	9C56	8818	蠟		
3026	鯿	1	9C6A	9C2E	鯿		
3069	三	1	9CB7			三	
3161	壘	2	9D54	58DC	壘		
3179	鎔	1	9D66	93E5	鎔		
3199	骹	1	9D7A	9ABD	骹		
3770	个	1	8141	4E2A	个		「個」，取代。
3788	慙	1	8153	22871		殼 ㄉㄨㄞ 心	
3797	凭	1	815C	51ED	凭		
3799	刼	2	815E	5226	刼		
3802	劬	1	8161	529A	劬		
3803	飢	2	8162	4B54		食 ㄇ 追	
3825	疊	1	8178	53E0	疊		
3830	咤	2	817D	54A4	咤		
3833	噱	1	81A2	5611	噱		
3835	噍	4	81A4	564D	噍		
3836	喻	2	81A5	564F	喻		
3860	尠	1	81BD	5C20	尠		
3862	屈	1	81BF	5C49	屈		
3895	淑	2	81E0	6E86	淑		
3932	櫻	2	8246	6936	櫻		
3935	槩	1	8249	3BA3		既 ㄉㄨㄞ 木	
3937	檝	1	824B	6A9D	檝		
3942	毡	2	8250	6BE1	毡		「氈」，取代。
3970	焗	4	826C	714F	焗		
3974	牀	1	8270	7240	牀		
3980	獾	2	8276	7328	獾		
3981	珉	3	8277	73C9	珉		
3982	珣	2	8278	73F7	珣		
4002	瘡	1	82AE	7607	瘡		

編號	造字	字頻	Big5	Unicode	WinXP	構字式	備註
4008	癯	1	82B4	766F	癯		
4030	秆	2	82CA	79C6	秆		
4034	穉	1	82CE	7A49	穉		
4035	穽	1	82CF	7A7D	穽		
4036	窰	6	82D0	7AB0	窰		
4041	豎	2	82D5	7AEA	豎		
4045	箒	2	82D9	7B92	箒		
4058	糲	1	82E6	7CC9	糲		
4065	紮	6	82ED	7D25	紮		
4067	綉	2	82EF	7D89	綉		
4069	綑	1	82F1	7DB3	綑		
4085	耻	1	8342	803B	耻		
4091	皐	4	8348	81EF	皐		
4102	苣	1	8353	82E2	苣		
4122	蛭	19	8367	87D5	蛭		
4124	蝨	1	8369	8771	蝨		
4130	衅	1	836F	8845	衅		
4167	賁	2	83B6	8CEB	賁		
4169	趲	1	83B8	8DA6	趲		
4172	屨	1	83BB			屨 𡳶 足	
4176	輒	4	83BF	8F19	輒		
4184	迹	15	83C7	8FF9	迹		
4190	邳	1	83CD	90C4	邳		
4208	磧	2	83DF	9767	磧		
4210	揄	1	83E1	2468E		𠂇 𡳶 俞	
4223	顙	1	83EE	984B	顙		
4250	鰭	1	844A	9C0C	鰭		
4254	鴟	1	844E	9D7E	鴟		
4265	鼃	2	8459	9F03	鼃		
4270	𩺰	1	845E	9F45	𩺰		
4275	𠄎	2	8463	4E49	𠄎		「義」，取代。
4291	縻	2	8473	431D		纟 𡳶 禁	

編號	造字	字頻	Big5	Unicode	WinXP	構字式	備註
4349	蟻	1	84CF	878A	蟻		「蟻」，錯字。
4446	龔	1	8571	9F47	龔		
4456	舫	1	857B	8684	舫		
4461	勛	1	85A2	52D4	勛		
4498	鵠	1	85C7	4CD5		孚 𪛗 鳥	
4524	戎	2	85E1			王 𪛗 戎	「王 𪛗 戎」，錯字。
4528	嶠	1	85E5	5D60	嶠		
4536	𪛗	2	85ED	3F10		扁 𪛗 瓜	
4558	厩	3	8644	53A9	厩		Unicode 字型呈現差異。
4575	縊	1	8655	7E7F	縊		
4592	獾	5	8666	247E4		犴 𪛗 軍	
4735	壠	1	8758	58E0	壠		
4767	懷	9	8778	61B9	懷		
4785	贗	1	87AC	8D0B	贗		
4793	祿	1	87B4	7958	祿		Unicode 字型呈現差異。
4797	堡	1	87B8	57A1	堡		
4803	脬	3	87BE	816D	脬		
4823	窑	1	87D2	7A91	窑		
4846	愷	6	87E9	398E		忄 𪛗 畫	
4914	杌	3	886E	233D5		木 𪛗 朮	
4998	𪛗	1	88E4	444B		月 𪛗 襄	
5025		2	8940			千 𪛗 干	「𪛗」，取代。
5033	𪛗	4	8948	8492	𪛗		「朮」，重碼部件，不轉碼。
5278	𪛗	1	8AC2	44B7		𪛗 𪛗 舌	
5281	𪛗	3	8AC5	994A	𪛗		
5318	𪛗	1	8AEA	4516		𪛗 𪛗 陵	
5335	𪛗	5	8AFB	459F		亡 𪛗 虫	
5351	𪛗	30	8B4C	80AC	𪛗		「𪛗」，重碼造字，手動取代。 「𪛗」，重碼部件，不轉碼。
5376	𪛗	4	8B65			鹿 𪛗 京	
5396	𪛗	1	8B79	938B	𪛗		
5399	𪛗	1	8B7C	26CDD		𪛗	「𪛗」，重碼部件，不轉碼。

編號	造字	字頻	Big5	Unicode	WinXP	構字式	備註
5442	枿	2	8BC9	67BF	枿		
5564	𣎵	2	8CA6			𣎵 𣎵 木	
5579	𣎵	1	8CB5	5D08	𣎵		
5582	𣎵	2	8CB8	97B5	𣎵		
5587	𣎵	2	8CBD	6B58	𣎵		
5591	𣎵	2	8CC1	233CC		木 𣎵 巳	
5592	𣎵	1	8CC2	3B9A		𣎵	
5654	𣎵	1	8D41	7C4B	𣎵		
5719	𣎵	1	8DA4	4D25		鹿 𣎵 加	
5720	𣎵	1	8DA5	578D	𣎵		
5772	𣎵	1	8DD9	6FD9	𣎵		
5793	𣎵	1	8DEE			鹿 𣎵 旨	
5858	𣎵	1	C6D1			𣎵	
5860	无	2	C6D3			无	

附件二、《本草綱目》手動取代表

編號	造字	部件	檔案原文摘錄	手動取代結果
309	?		从木， <u> </u> 聲。易音陽， <u> </u> 音酉。	从木， <u>𣎵</u> 聲。易音陽， <u>𣎵</u> 音酉。
781	丿		<u>芳</u> <u>𣎵</u>	<u>芬</u>
5351	𣎵	𣎵	蝕 <u>𣎵</u>	蝕 <u>𣎵</u>
			燕趙之間謂之蝕[五] <u>𣎵</u> 。	燕趙之間謂之蝕[五] <u>𣎵</u> 。
			<u>𣎵</u> 即疣子	<u>𣎵</u> 即疣子
			今人病 <u>𣎵</u> 者	今人病 <u>𣎵</u> 者