

# 《丹溪醫集》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  
2008/3/17 黃榮順 製作

一、《丹溪醫集》一書(丹溪醫集.xml)使用舊版造字 74 個，字頻 141 次，詳如附件一。這 74 個造字中，58 個可轉成 Windows XP 可顯示的字，字頻 120 次；另外 16 個字必須轉成構字式，字頻 21 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，舊版漢籍造字用為構字部件，手動取代：txt 檔使用的造字到 xml 檔中已被修改完畢，但 xml 檔中其他的構字式碰巧使用了該造字，因有字頻差異，故須手動取代。如編號 785 的「ノ」，在 txt 檔中的 2 次已被修改完畢，而在 xml 檔新出現的構字式中又使用了 1 次「ノ」，因頻次不同，故須手動取代。
- 2、標記問題，新版漢籍表格內文字以圖版替代：新版漢籍的表格一律以圖版方式呈現，使得舊版漢籍表格內的文字在新版漢籍中不再以內文方式呈現。故若舊版漢籍表格內的文字出現缺字，因新版漢籍中已改用圖版呈現，而造成 txt 檔與 xml 檔缺字頻次不符的情形。如編號 5448 的「髑」字，txt 檔出現 2 次，而 xml 檔只有 1 次，txt 檔多出的 1 次出現在原書 521 頁的表格中，由於 xml 檔已將表格用圖版取代，所以手動將 xml 檔頻次改為「髑」字。
- 3、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 982 的「𪛗」係「胸」字之異體，故以「胸」字取代。又如編號 2197 的「𪛗」係「鐵」字之簡體，故以「鐵」字取代。

- 4、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 4145 的「羈」字，Unicode 字型呈現為「羈」，實際上仍為同一字。
- 5、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 5749 的「片𪛗庸」。

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 1041 的「𪛗」字，Unicode 編碼為 46A1，由於 Windows XP 並不支援，仍須改用構字式「角𪛗思」。

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：舊漢籍造字。
- 丙、取代：使用字形與造字欄不同時，於此欄列出。
- 丁、檔案原文摘錄：摘錄檔案(丹溪醫集.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。

五、「丹溪醫集.xml」缺字外發現之錯字共 1 處，修正情形詳列附件三，附件三「丹溪醫集.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(丹溪醫集.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、「丹溪醫集.xml」造字分析表

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
489	隶	1	1	FD51			隶	
750	𠂔	1	1	FEDB			𠂔	
785	ノ	2	1	FEFE			ノ	ノ，重碼問題，舊版漢籍造字用為構字部件，手動取代。 𠂔，該構字式又為「𠂔」之異體，取代之。
912	鷓	1	1	8EE0	9DC0	鷓		
982	𦉳	1	1	8F67	80F7	𦉳		胸，異體字問題。
1041	𦉳	1	1	8FC4	46A1		角𠂔思	
1145	稭	1	1	906D	7A2C	稭		
1146	蕘	2	2	906E	7506	蕘		
1278	碍	2	2	9155	788D	碍		
1321	確	1	1	91A2	7936	確		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
1555	緡	1	1	92EF	7DCD	緡		
1905	裊	1	1	9554	88E9	裊		
2042	疏	3	3	9640	47FD		足厶充	疏，異體字問題。
2046	踪	2	2	9644	8E2A	踪		
2142	裊	2	2	96C6	7758	裊		
2147	蟪	1	1	96CB	87EE	蟪		
2168	醅	1	1	96E0	91BB	醅		
2197	鉄	2	2	96FD	9244	鉄		鐵，異體字問題。
2305	鎖	1	1	97CC	93BB	鎖		
2409	関	4	4	9875	95A7	関		
2524	饒	1	1	994B	994D	饒		
2649		2	2	99EA	28891		酉厶兼	鹵，異體字問題。
2719	函	1	1	9A71	51FE	函		
2885	呌	1	1	9B7A	544C	呌		
3021	卜	2	2	9C65			卜	
3770	个	3	3	8141	4E2A	个		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
3858	寃	1	1	81BB	5BC3	寃		
3864	峯	1	1	81C1	5CEF	峯		
3927	柏	6	6	8241	6855	柏		
3931	椀	2	2	8245	6900	椀		
3937	檝	1	1	824B	6A9D	檝		
3940	爵	1	1	824E	6B1D	爵		
3944	氷	1	1	8252	6C37	氷		
3945		1	1	8253			汙 𠄎义	汙，異體字問題。
3951	湏	1	1	8259	6E4F	湏		
3957	凜	2	2	825F	6F9F	凜		
3967	炳	3	3	8269	712B	炳		
3992	嘗	1	1	82A4	751E	嘗		
4000	疴	1	1	82AC	75B4	疴		
4041	豎	3	3	82D5	7AEA	豎		
4042	笋	3	3	82D6	7B0B	笋		
4093	舐	1	1	834A	445B		舌 𠄎 舐	舐，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4094	館	3	3	834B	8218	館		
4095	舩	1	1	834C	8229	舩		
4140	衹	1	1	8379	88B5	衹		
4145	羈	2	2	837E	898A	羈		Unicode 字型差異。
4146	覩	2	2	83A1	89A9	覩		
4178	輓	7	7	83C1	8F2D	輓		
4190	邈	1	1	83CD	90C4	邈		
4226	淪	1	1	83F1	98E1	淪		
4278	亘	1	1	8466	4E98	亘		
4447	捏	7	7	8572	63D1	捏		
4468	鱸	5	5	85A9	9F44	鱸		
4515	呪	2	2	85D8	546A	呪		
4630		1	1	86AE			𠄎匕示欠𠄎	款，異體字問題。
4649	蕒	1	1	86C1	8552	蕒		
4668	噐	5	5	86D4	5650	噐		
4803	腭	1	1	87BE	816D	腭		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4894		1	1	885A			宀 ㇀ 心 ㇀ 丹	窟，異體字問題。
4910		1	1	886A			氵 ㇀ 畚	淫，異體字問題。
4955	龜	2	2	88B9	9F9C		龜	
5008	幪	1	1	88EE	5E5E	幪		
5032	焮	1	1	8947	3B0E		焮	
5074	𠂔	1	1	8971	5C2B	𠂔		
5098	寤	1	1	89AB	7AB9	寤		
5241	鷺	1	1	8A7B	9D5E	鷺		
5335	𧈧	1	1	8AFB	459F		亡 ㇀ 虫	虻，異體字問題。
5440	昧	1	1	8BC7	771C	昧		Unicode 字型差異。
5448	𦉳	2	1	8BCF	9AC3	𦉳		𦉳，校對問題，xml 檔將表格替換為圖檔。
5571	樛	1	1	8CAD	6A45	樛		
5693	𪗇	5	5	8D68	9E7B	𪗇		
5742	澁	11	11	8DBB	6F81	澁		
5749		1	1	8DC2			片 ㇀ 庸	待造字。
5768	潤	1	1	8DD5	6FF6	潤		

附件二、「丹溪醫集.xml」手動取代表

編號	造字	取代	檔案原文摘錄	手動取代結果
785	ノ	肩	心 <u>厶局</u> 融廓	心 <u>肩</u> 融廓
5448	髑		絕骨、環跳、肩 <u>音</u>	絕骨、環跳、肩 <u>髑</u>

附件三、「丹溪醫集.xml」缺字外錯字取代表

檔案原文摘錄	手動取代結果
心 厶局融 <u>廡</u>	心 厶局融 <u>廡</u>