

《東垣醫集》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

2008/3/26 陳建安 製作

一、《東垣醫集》一書(東垣醫集.xml)使用舊版造字 50 個，字頻 103 次，詳如附件一。這 50 個造字中，43 個可轉成 Windows XP 能顯示的字，字頻 82 次；另外 7 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、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 1149 的「豕𪛗生」係「豨」字之異體，故以「豨」字取代。又編號 1513 的「糧」字，係「糧」字之簡體字，亦以正體的「糧」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 4145 的「羈」字，Unicode 字型呈現為「羈」，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 5335 的「𧈧」字，Unicode 編碼為 459F，由於 Windows XP 並不支援，仍須使用構字式「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

附件一、《東垣醫集》造字分析表

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
1149		1	1	9071			豕𡗗生	豨，異體字問題。
1513	糧	1	1	92C5	7CAE	糧		糧，異體字問題。
1981	讐	1	1	95C2	8B90	讐		
2325	瞶	2	2	97E0	77A4	瞶		
3071	三	1	1	9CB9			三	
3230	芦	1	1	9DBB	82A6	芦		蘆，異體字問題。
3796	決	18	18	815B	51B3	決		決，異體字問題。
3817	廁	2	2	8170	53A0	廁		廁，異體字問題。
3822	叁	1	1	8175	53C1	叁		參，異體字問題。
3824	叙	2	2	8177	53D9	叙		敘，異體字問題。
3826		2	2	8179	53F6	叶		葉，異體字問題。
3951	湏	1	1	8259	6E4F	湏		
3967	炳	2	2	8269	712B	炳		
3968	煅	2	2	826A	7145	煅		
4002	瘡	1	1	82AE	7607	瘡		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4012	𪗇	1	1	82B8	76B7	𪗇		
4013	𪗈	1	1	82B9	76CC	𪗈		
4044	𪗉	1	1	82D8	7B6F	𪗉		
4068	𪗊	1	1	82F0	7DAB	𪗊		線，異體字問題。
4085	𪗋	1	1	8342	803B	𪗋		
4091	𪗌	1	1	8348	81EF	𪗌		
4093	𪗍	2	2	834A	445B		舌𪗍氏	𪗍，異體字問題。
4100	𪗎	2	2	8351	82FD	𪗎		
4118	𪗏	4	4	8363	8602	𪗏		
4124	𪗐	2	2	8369	8771	𪗐		
4131	𪗑	3	3	8370	8842	𪗑		
4145	𪗒	1	1	837E	898A	𪗒		Unicode 呈現差異。
4173	𪗓	1	1	83BC	8EAD	𪗓		
4176	𪗔	1	1	83BF	8F19	𪗔		
4184	迹	2	2	83C7	8FF9	迹		
4186	𪗕	1	1	83C9	9061	𪗕		
4190	𪗖	2	2	83CD	90C4	𪗖		
4223	𪗗	1	1	83EE	984B	𪗗		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4226	飡	2	2	83F1	98E1	飡		
4240	髥	1	1	8440	9AF4	髥		
4266	鼃	1	1	845A	9F02	鼃		
4286	从	3	3	846E	4ECE	从		從，異體字問題。
4345		12	12	84CB			口𠂔敕	漱，異體字問題。
5119	博	2	2	89C0	613D	博		
5240	擗	2	2	8A7A	6481	擗		
5335	𧈧	3	3	8AFB	459F		亡𠂔虫	
5353	役	1	1	8B4E	4F07	役		役，異體字問題。
5439	栢	1	1	8BC6	6822	栢		柏，異體字問題。
5558	龕	2	2	8C7E	9E84	龕		
5568	开	1	1	8CAA	5F00	开		開，異體字問題。
5621	汨	2	2	8CDF	6CEA	汨		淚，異體字問題。
5622	邛	1	1	8CE0	48B5		云𠂔卩	
5706	敍	1	1	8D75	654D	敍		敘，異體字問題。
5768	濶	1	1	8DD5	6FF6	濶		闊，異體字問題。
5803	𧈧	1	1	8DF8	4ABC		出𠂔頁	