

《湘山野錄》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
2008/6/20 葉冠孜 製作

一、《湘山野錄》一書(湘山野錄.xml)使用舊版造字 95 個，字頻 360 次，詳如附件一。這 95 個造字中，76 個可轉成 Windows XP 能顯示的字，字頻 276 次；另外 19 個字必須轉成構字式，字頻 84 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、標記問題：標記問題主要是漢籍電子文獻在舊轉新的過程中，採用了新的標記語言，而在修改標記時，對於某些標題句做了增刪的動作，因為這些標題句的內容包含了缺字，所以這些更改標記的動作造成 txt 檔與 xml 檔缺字頻次不符的情形。如編號 4507 的「吳」字，xml 檔在 91 頁少了一個標題，因而造成頻次不同。

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 4134 的「衛」係「衛」字之異體，故以「衛」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 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 4673 的「茱」字，Unicode

編碼為 44C7，由於 Windows XP 並不支援，仍須使用構字式「𠄎赤」。

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、txt 檔與 xml 檔頻次不同且造字分析表「備註」欄有備註問題情況者，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：實際使用字形與舊漢籍造字不同時，於此欄列出。

丁、檔案原文摘錄：摘錄檔案(湘山野錄.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。文句後方標示(全部取代)者，表示執行全部取代的動作，不一一列舉所有原文。

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。文句後方標示(全部取代)者，表示執行全部取代的動作，不一一列舉所有修改後結果。

附件一、《湘山野錄》造字分析表

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
10	𠄎	1	1	FA49	8FA4	𠄎		
296	𠄎	3	3	FBEC			𠄎	
297	𠄎	4	4	FBED			𠄎	

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
298	△	1	1	FBEE			△	
299	形	15	15	FBEF			形	
300	◻	15	15	FBF0			◻	
379	支	1	1	FCA2			支	
410	昇	1	1	FCC1			昇	
662	聿	9	9	FE61			聿	
765	冫	5	5	FEEA			冫	
832	沱	1	1	8E6E	6CB2	沱		
887	鷺	1	1	8EC7	9E0E	鷺		
893	峩	2	2	8ECD	5CE9	峩		峨，異體字問題。
972	犇	3	3	8F5D	7287	犇		
1039	稟	3	3	8FC2	69C0	稟		
1067	碁	2	2	8FDE	68CA	碁		棋，異體字問題。
1133	瓌	3	3	9061	74CC	瓌		
1200	眦	1	1	90C6			目△比	
1288	畝	2	2	915F	342D		畝	

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
1565	繇	1	1	92F9	7DDC	繇		
1893	袞	8	8	9548	886E	袞		
1944	簪	1	1	957B	8A67	簪		
1981	讐	1	1	95C2	8B90	讐		
2021	趲	2	2	95EA	8D82	趲		
2046	踪	3	3	9644	8E2A	踪		
2157	酌	1	1	96D5	9167	酌		
2490	頷	1	1	98E8			名△頁	
2826	巡	1	1	9AFE	5EF5	巡		巡，異體字問題。
2885	呌	2	2	9B7A	544C	呌		
3789	冲	3	3	8154	51B2	冲		冲，異體字問題。
3792	凉	6	6	8157	51C9	凉		凉，異體字問題。
3797	凭	1	1	815C	51ED	凭		
3805	勅	1	1	8164	52C5	勅		
3814	却	10	10	816D	5374	却		卻，異體字問題。
3817	廁	1	1	8170	53A0	廁		廁，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
3829	咏	1	1	817C	548F	咏		
3846	壻	1	1	81AF	58FB	壻		Unicode 呈現差異。
3858	寃	1	1	81BB	5BC3	寃		寃，異體字問題。
3860	尠	1	1	81BD	5C20	尠		
3876	廻	1	1	81CD	5EFB	廻		廻，異體字問題。
3883	徧	1	1	81D4	5FA7	徧		
3884	忽	6	6	81D5	6031	忽		
3886	悞	1	1	81D7	609E	悞		誤，異體字問題。
3891	慙	3	3	81DC	6159	慙		慙，異體字問題。
3916	敷	1	1	81F5	656D	敷		
3937	檝	3	3	824B	6A9D	檝		
3974	牀	6	6	8270	7240	牀		
3996	疎	4	4	82A8	758E	疎		
4014	蓋	36	36	82BA	8462	蓋		蓋，異體字問題。
4016	着	2	2	82BC	7740	着		著，異體字問題。
4039	竚	1	1	82D3	7ADA	竚		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4040	竝	1	1	82D4	7ADD	竝		
4056	葬	2	2	82E4	7CA6	葬		
4080	羣	9	9	82FC	7FA3	羣		群，異體字問題。
4095	舩	1	1	834C	8229	舩		
4109	葛	1	1	835A	84AD	葛		
4118	藥	8	8	8363	8602	藥		
4132		13	13	8371			血𠄎𠄎	眾，異體字問題。
4134	衛	1	1	8373	885E	衛		衛，異體字問題。
4146	覩	4	4	83A1	89A9	覩		
4158	譌	2	2	83AD	8B4C	譌		Unicode 呈現差異。
4167	賈	1	1	83B6	8CEB	賈		
4184	迹	9	9	83C7	8FF9	迹		跡，異體字問題。
4195	鉢	1	1	83D2	9262	鉢		
4220	穎	1	1	83EB	9834	穎		穎，異體字問題。
4242	鬪	2	2	8442	9B2D	鬪		Unicode 呈現差異。
4309	劫	2	2	84A7	523C	劫		劫，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4455	劒	3	3	857A			僉𠄎刀	劍，異體字問題。
4460	勅	8	8	85A1	52D1	勅		
4507	吳	42	41	85D0	5434	吳		吳，標記問題，以標準字取代。
4515	呪	1	1	85D8	546A	呪		
4546	鴟	1	1	85F7	9D76	鴟		
4558	廐	2	2	8644	53A9	廐		廐，異體字問題。
4603	憇	4	4	8671	6187	憇		
4673	茛	1	1	86D9	44C7		艹𠄎赤	
4750	卅	1	1	8767			卅	
4805	瀆	1	1	87C0	51DF	瀆		瀆，異體字問題。
4817	寶	7	7	87CC	5BF3	寶		寶，異體字問題。
4878	宀	1	1	884A	5B82	宀		宀，異體字問題。
4904	尅	1	1	8864	5C05	尅		尅，異體字問題。
4940	韵	1	1	88AA	97F5	韵		
4957	崑	1	1	88BB	5D53	崑		
5008	幪	1	1	88EE	5E5E	幪		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
5028	昇	3	3	8943	662A	昇		
5031	闌	1	1	8946	95B4	闌		闌，異體字問題。
5176		3	3	89F9			冫厶冗	沉，異體字問題。
5225	卹	4	4	8A6B	460F		血厶卩	卹，異體字問題。
5290	携	23	23	8ACE	643A	携		
5302	昏	1	1	8ADA	662C	昏		昏，異體字問題。
5447	躡	1	1	8BCE	8E83	躡		
5467	琴	1	1	8BE2	7439	琴		
5559	欸	2	2	8CA1	6B35	欸		
5583	殮	1	1	8CB9	98F1	殮		殮，異體字問題。
5588	恠	1	1	8CBE	6061	恠		
5591	杞	1	1	8CC1	233CC		木厶巳	杞，異體字問題。

附件二、《湘山野錄》手動取代表

編號	造字	取代	檔案原文摘錄	手動取代結果
4507	吳	吳	吳	吳（全部取代）