

《巢林筆談》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
2008/8/11 葉冠孜 製作

一、《巢林筆談》一書(巢林筆談.xml)使用舊版造字 135 個，字頻 622 次，詳如附件一。這 135 個造字中，101 個可轉成 Windows XP 能顯示的字，字頻 452 次；另外 34 個字必須轉成構字式，字頻 170 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 4132 的「血~~𠂇~~𠂇」係「眾」字之異體，故以「眾」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 3846 的「壻」字，Unicode 字型呈現為「壻」，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 1616 的「運」字，Unicode 編碼為 284E6，由於 Windows XP 並不支援，仍須使用構字式「~~辶~~里」。

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、《巢林筆談》造字分析表

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
296	𠄎	19	19	FBEC			𠄎	
297	𠄏	24	24	FBED			𠄏	
298	𠄐	13	13	FBEE			𠄐	
299	𠄑	12	12	FBEF			𠄑	
300	𠄒	12	12	FBF0			𠄒	
337	𠄓	1	1	FC56			𠄓	
379	支	2	2	FCA2			支	
658	𠄔	9	9	FE5D			𠄔	
685	𠄕	1	1	FE78			𠄕	
750	𠄖	1	1	FEDB			𠄖	
752	𠄗	13	13	FEDD			𠄗	

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
771	攵	1	1	FEF0			攵	
773	厶	1	1	FEF2			厶	
777	宀	2	2	FEF6			宀	
780	彳	1	1	FEF9			彳	
784	亻	1	1	FEFD			亻	
785	ノ	1	1	FEFE			ノ	
1076	高	1	1	8FE7	5368	高		Unicode 呈現差異。
1433	宀	1	1	9253	53DC	宀		
1544	絀	1	1	92E4	7D59	絀		
1616	逕	1	1	936D	284E6		逕△里	
1664	腸	2	2	93BF	8193	腸		
1971	宀	2	2	95B8	5B8D	宀		
2021	趣	1	1	95EA	8D82	趣		
2046	踪	3	3	9644	8E2A	踪		
2143	火	1	1	96C7			火	
3247	鎬	1	1	9DCC	9424	鎬		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
3770	个	1	1	8141	4E2A	个		
3789	冲	2	2	8154	51B2	冲		
3791	况	2	2	8156	51B5	况		
3796	决	2	2	815B	51B3	决		
3804	効	1	1	8163	52B9	効		
3814	却	23	23	816D	5374	却		
3817	厠	4	4	8170	53A0	厠		
3824	叙	4	4	8177	53D9	叙		
3826	叶	1	1	8179	53F6	叶		
3829	咏	10	10	817C	548F	咏		
3831	喆	2	2	817E	5586	喆		
3834	嗶	2	2	81A3	5637	嗶		
3846	壻	7	7	81AF	58FB	壻		Unicode 呈現差異。
3852	媼	2	2	81B5	5A63	媼		
3858	寃	19	19	81BB	5BC3	寃		
3876	廻	5	5	81CD	5EFB	廻		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
3883	徧	2	2	81D4	5FA7	徧		
3891	慙	6	6	81DC	6159	慙		
3899	担	2	2	81E4	62C5	担		
3907	搯	1	1	81EC	644F	搯		
3921	皙	2	2	81FA	6673	皙		
3935	槩	1	1	8249	69E9	槩		Unicode 呈現差異。
3950	浜	1	1	8258	6D5C	浜		
3966	烟	33	33	8268	70DF	烟		
3969	煊	3	3	826B	714A	煊		
3974	牀	9	9	8270	7240	牀		
3975	牕	8	8	8271	7255	牕		Unicode 呈現差異。
3977	狗	5	5	8273	72E5	狗		
3978	猪	1	1	8274	732A	猪		Unicode 呈現差異。
3991	攜	2	2	82A3	64D5	攜		
3996	疎	10	10	82A8	758E	疎		
4000	疴	1	1	82AC	75B4	疴		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4016	着	13	13	82BC	7740	着		Unicode 呈現差異。
4028	稟	2	2	82C8	7980	稟		
4036	窰	1	1	82D0	7AB0	窰		
4038	竈	5	5	82D2	7AC8	竈		
4041	豎	1	1	82D5	7AEA	豎		
4067	綉	1	1	82EF	7D89	綉		
4069	綑	3	3	82F1	7DB3	綑		
4077	罇	1	1	82F9	7F47	罇		
4080	羣	10	10	82FC	7FA3	羣		
4081	翱	3	3	82FD	7FFA	翱		
4085	耻	1	1	8342	803B	耻		
4090	脚	2	2	8347	811A	脚		
4091	臯	1	1	8348	81EF	臯		
4094	館	4	4	834B	8218	館		
4118	藥	2	2	8363	8602	藥		
4132		27	27	8371			血𠂇𠂇	眾，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4134	衛	10	10	8373	885E	衛		
4144	霸	1	1	837D	8987	霸		Unicode 呈現差異。
4145	羈	2	2	837E	898A	羈		Unicode 呈現差異。
4146	覩	1	1	83A1	89A9	覩		
4172	屨	1	1	83BB	8E77	屨		Unicode 呈現差異。
4174	躑	1	1	83BD	8EB6	躑		
4184	迹	2	2	83C7	8FF9	迹		
4188	邨	1	1	83CB	90A8	邨		
4193	鈎	3	3	83D0	920E	鈎		
4195	鉢	1	1	83D2	9262	鉢		
4203	隣	7	7	83DA	96A3	隣		
4216	鞮	1	1	83E7	97C8	鞮		
4221	頰	1	1	83EC	9833	頰		
4242	鬪	1	1	8442	9B2D	鬪		Unicode 呈現差異。
4256	鷄	1	1	8450	9DC4	鷄		
4270	鼻	1	1	845E	9F45	鼻		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4278	亘	2	2	8466	4E98	亘		
4307	罵	3	3	84A5	99E1	罵		
4309	刼	2	2	84A7	523C	刼		
4442	潛	1	1	856D	6F98	潛		Unicode 呈現差異。
4448		1	1	8573			𦉳 纟 厂 墨 𠂇	纏，異體字問題。
4482	厯	1	1	85B7	53AA	厯		
4496	段	8	8	85C5	53DA	段		
4504	呬	1	1	85CD	541A	呬		
4507	吴	90	90	85D0	5434	吴		
4542	塿	4	4	85F3	21484		土 厶 稟	
4603	憇	2	2	8671	6187	憇		
4702	場	6	6	86F6	5872	場		
4750	卅	2	2	8767			卅	
4768	妬	18	18	8779	59AC	妬		
4785	贗	2	2	87AC	8D0B	贗		
4805	瀆	1	1	87C0	51DF	瀆		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4839	昱	1	1	87E2	6637	昱		
4840	昇	1	1	87E3			西 厶 升	
4841	震	1	1	87E4	290D9		雷 厶 大	
4842	霽	1	1	87E5	29166		雨 厶 單	
4907	橐	1	1	8867	69D6	橐		
4940	韵	5	5	88AA	97F5	韵		
5019	畧	6	6	88F9			畧	
5131	窻	1	1	89CC	7ABB	窻		Unicode 呈現差異。
5173	漱	1	1	89F6	6F44	漱		
5194		2	2	8A4C			扌 畺 且 口	擅，異體字問題。
5195		1	1	8A4D			木 畺 且 口	檀，異體字問題。
5221		2	2	8A67			白 米 丂 口	粵，異體字問題。
5230		2	2	8A70			土 畺 且 口	壇，異體字問題。
5299	戟	1	1	8AD7	39B8		卓 厶 戈	
5302	昏	3	3	8ADA	662C	昏		
5338	暎	5	5	8AFE	668E	暎		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
5439	栢	6	6	8BC6	6822	栢		
5541	蠹	6	6	8C6D	8827	蠹		
5551	賈	1	1	8C77	8CF7	賈		
5556	隸	1	1	8C7C	96B7	隸		
5559	欸	1	1	8CA1	6B35	欸		
5587	歛	2	2	8CBD	6B58	歛		
5591	杞	2	2	8CC1	233CC		木厶巳	
5592	桌	1	1	8CC2	3B9A		桌	
5682	蠹	1	1	8D5D	881A	蠹		
5702	厨	1	1	8D71	53A8	厨		
5704	穉	1	1	8D73	6C0A	穉		
5706	敍	7	7	8D75	654D	敍		