

《歸田錄》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
2008/5/13 葉冠孜 製作

一、《歸田錄》一書(歸田錄.xml)使用舊版造字 61 個，字頻 143 次，詳如附件一。這 61 個造字中，52 個可轉成 Windows XP 能顯示的字，字頻 116 次；另外 9 個字必須轉成構字式，字頻 27 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 3935 的「槩」係「概」字之異體，故以「概」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 868 的「休」字，Unicode 編碼為 201F1，由於 Windows XP 並不支援，仍須使用構字式「亻𠂔末」。

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、《歸田錄》造字分析表

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
296	厶	4	4	FBEC			厶	
297	厶	4	4	FBED			厶	
675	食	2	2	FE6E			食	
777	宀	2	2	FEF6			宀	
783	宀	2	2	FEFC			宀	
868	侏	1	1	8EB4	201F1		亻 厶 末	
1024	虵	1	1	8FB3	8675	虵		
1039	稟	3	3	8FC2	69C0	稟		
1053	傲	1	1	8FD0	6160	傲		
1146	甕	1	1	906E	7506	甕		
1331	嶠	2	2	91AC	3813		山 厶 蒙	
1893	袞	2	2	9548	886E	袞		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
2021	趁	1	1	95EA	8D82	趁		
2151	𠃉	2	2	96CF	961D	𠃉		
3791	况	2	2	8156	51B5	况		况，異體字問題
3792	涼	1	1	8157	51C9	涼		涼，異體字問題
3814	却	3	3	816D	5374	却		卻，異體字問題
3817	廁	1	1	8170	53A0	廁		廁，異體字問題
3824	叙	1	1	8177	53D9	叙		敘，異體字問題
3826	叶	1	1	8179	53F6	叶		
3829	咏	1	1	817C	548F	咏		
3846	壻	5	5	81AF	58FB	壻		Unicode 呈現差異
3883	徧	1	1	81D4	5FA7	徧		
3935	槩	2	2	8249	69E9	槩		概，異體字問題
3936	棹	1	1	824A	6A70	棹		
3949	涖	1	1	8257	6D96	涖		
3979	猫	2	2	8275	732B	猫		貓，異體字問題
3996	疎	3	3	82A8	758E	疎		
4012	鼓	1	1	82B8	76B7	鼓		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4016	着	8	8	82BC	7740	着		著，異體字問題
4028	稟	1	1	82C8	7980	稟		稟，異體字問題
4044	筋	2	2	82D8	7B6F	筋		
4067	綉	1	1	82EF	7D89	綉		繡，異體字問題
4077	罇	3	3	82F9	7F47	罇		
4080	羣	17	17	82FC	7FA3	羣		群，異體字問題
4090	脚	1	1	8347	811A	脚		腳，異體字問題
4132		9	9	8371			血𠄎𠄎	眾，異體字問題
4134	衛	9	9	8373	885E	衛		衛，異體字問題
4176	輒	4	4	83BF	8F19	輒		
4178	輓	1	1	83C1	8F2D	輓		
4184	迹	1	1	83C7	8FF9	迹		
4193	鈎	1	1	83D0	920E	鈎		
4203	隣	1	1	83DA	96A3	隣		鄰，異體字問題
4219	頰	1	1	83EA	982E	頰		
4240	髯	1	1	8440	9AF4	髯		
4247	鮎	2	2	8447	9B8E	鮎		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4286	从	1	1	846E	4ECE	从		
4287	織	1	1	846F	7E8E	織		織，異體字問題
4460	勅	4	4	85A1	52D1	勅		敕，異體字問題
4507	吳	4	4	85D0	5434	吳		吳，異體字問題
4904	尅	1	1	8864	5C05	尅		
5019	畧	1	1	88F9			畧	略，異體字問題
5179	駟	1	1	89FC	99C8	駟		
5289	愨	1	1	8ACD	6164	愨		愨，異體字問題
5290	携	1	1	8ACE	643A	携		攜，異體字問題
5517	楹	1	1	8C55	6985	楹		Unicode 呈現差異
5542	鑠	1	1	8C6E	93C1	鑠		
5549	藁	2	2	8C75	85F3	藁		
5704	羶	1	1	8D73	6C0A	羶		
5706	敍	5	5	8D75	654D	敍		敍，異體字問題
5802	標	1	1	8DF7	5E56	標		