

《舊聞證誤》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
2008/6/24 陳維君 製作

一、《舊聞證誤》一書(舊聞證誤.xml)使用舊版造字 33 個，字頻 88 次，詳如附件一。這 33 個造字中，27 個可轉成 Windows XP 能顯示的字，字頻 71 次；另外 6 個字必須轉成構字式，字頻 17 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「血𩇑𩇑」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 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 1331 的「嶸」字，Unicode 編碼為 3813，由於 Windows XP 並不支援，仍須使用構字式「山𩇑蒙」。

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、《舊聞證誤》造字分析表

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
297	厶	1	1	FBED			厶	
979	蕘	1	1	8F64	56CF	蕘		
1226	罰	2	2	90E0	7F78	罰		
1331	嶠	1	1	91AC	3813		山厶蒙	
1565	繇	1	1	92F9	7DDC	繇		
1572		1	1	9341			𠂇革口王𠂇	革厶呈，異體字問題。
1893	袞	3	3	9548	886E	袞		袞，異體字問題。
2565	霏	1	1	9974	9714	霏		
2821	悞	1	1	9AF9	394F		忄厶典	
3813	邛	1	1	816C	536D	邛		
3852	嫵	1	1	81B5	5A63	嫵		
3883	徧	1	1	81D4	5FA7	徧		
3935	槩	1	1	8249	69E9	槩		概，異體字問題。
3971	熇	1	1	826D	71BA	熇		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
3974	牀	1	1	8270	7240	牀		
4027	裨	1	1	82C7	7986	裨		裨，異體字問題。
4078		1	1	82FA			巳𠂇卅	异，異體字問題。
4080	羣	7	7	82FC	7FA3	羣		群，異體字問題。
4132		12	12	8371			血𠂇𠂇	眾，異體字問題。
4134	衛	3	3	8373	885E	衛		衛，異體字問題。
4146	覩	2	2	83A1	89A9	覩		
4158	譌	2	2	83AD	8B4C	譌		Unicode 呈現差異。
4184	迹	1	1	83C7	8FF9	迹		
4460	勅	12	12	85A1	52D1	勅		
4461	勅	5	5	85A2	52D4	勅		
4469	勻	1	1	85AA	5304	勻		
4507	吳	13	13	85D0	5434	吳		吳，異體字問題。
4629	埤	1	1	86AD	577F	埤		
5028	昇	1	1	8943	662A	昇		
5305	晒	3	3	8ADD	661E	晒		
5588	恠	1	1	8CBE	6061	恠		
5720	埤	1	1	8DA5	578D	埤		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
5761	慮	3	3	8DCE	61EC	慮		