

《瘟疫論》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
2008/3/19 黃榮順 製作

一、《瘟疫論》一書(瘟疫論.xml)使用舊版造字 38 個，字頻 74 次，詳如附件一。這 38 個造字中，30 個可轉成 Windows XP 可顯示的字，字頻 59 次；另外 8 個字必須轉成構字式，字頻 15 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 4080 的「羣」係「群」字之異體，故以「群」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。
- 3、待造字：Unicode 及漢字構形資料庫皆未收錄的舊漢籍造字，正在等待補造字中，所以「造字」欄空白無法看到字形。如編號 5749 的「片𠂔庸」。

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。所以如造字編號 3168 的「瘡」字，Unicode 編碼為 24EA5，由於 Windows XP 並不支援，仍須改用構字式「疒𠂔答」。

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、「瘟疫論.xml」造字分析表

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
493	胃	1	1	FD55	5190	胃		
2434		1	1	98B0			𠂔 𠂔 道	隨，異體字問題。
3168	瘡	2	2	9D5B	24EA5		疒 𠂔 答	
3789	冲	2	2	8154	51B2	冲		冲，異體字問題。
3792	凉	4	4	8157	51C9	凉		凉，異體字問題。
3796	决	1	1	815B	51B3	决		決，異體字問題。
3825	疊	1	1	8178	53E0	疊		疊，異體字問題。
3829	咏	1	1	817C	548F	咏		
3846	壻	1	1	81AF	58FB	壻		Unicode 字型差異。
3858	冤	1	1	81BB	5BC3	冤		
3886	悞	4	4	81D7	609E	悞		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
3929	黎	1	1	8243	68C3	黎		
3935	槩	5	5	8249	69E9	槩		概，異體字問題。
3940	藪	1	1	824E	6B1D	藪		
3968	煅	1	1	826A	7145	煅		
3974	牀	7	7	8270	7240	牀		
3996	疎	4	4	82A8	758E	疎		
4002	瘡	2	2	82AE	7607	瘡		
4004	瘻	1	1	82B0	763B	瘻		
4016	着	1	1	82BC	7740	着		著，異體字問題。
4038	竈	2	2	82D2	7AC8	竈		
4080	羣	1	1	82FC	7FA3	羣		群，異體字問題。
4131	𠂔	2	2	8370	8842	𠂔		
4184	迹	1	1	83C7	8FF9	迹		跡，異體字問題。
4193	鈎	1	1	83D0	920E	鈎		
4345		2	2	84CB			口𠂔敕	嗽，異體字問題。
4432		6	6	8563			𠂔𠂔厂墨𠂔	纏，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4643	徃	2	2	86BB	5F83	徃		
4767	懷	1	1	8778	61B9	懷		
4904	尅	1	1	8864	5C05	尅		
5019	畧	3	3	88F9	7567	畧		略，異體字問題。
5176		1	1	89F9			氵厶冗	沉，異體字問題。
5335	𧈧	1	1	8AFB	459F		亡𧈧虫	虻，異體字問題。
5342	曆	1	1	8B43			𠂔厂林日	曆，異體字問題。
5439	栢	2	2	8BC6	6822	栢		
5742	澁	3	3	8DBB	6F81	澁		
5749		1	1	8DC2			片厶庸	待造字。
5768	濶	1	1	8DD5	6FF6	濶		闊，異體字問題。