

《新序》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
2008/2/14 陳建安 製作

一、《新序》一書(新序.xml)使用舊版造字 69 個，字頻 165 次，詳如附件一。這 69 個造字中，58 個可轉成 Windows XP 能顯示的字，字頻 136 次；另外 11 個字必須轉成構字式，字頻 29 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、備註凡例：

- 1、校對問題，舊版漢籍錯字，可用程式全部取代：在舊版漢籍電子文獻中即存在的錯字，因校對時的疏漏而未更正，持續留存在新版漢籍電子文獻中；若該造字的所有頻次，皆屬於錯誤使用的錯字情形，可以用程式全部取代為正確字形。如編號 2449 的「名𪛗隹」字，原字為「雒」。
- 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 4252 的「魚𪛗眾」係「鰓」字之異體，故以「鰓」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 4558 的「厩」字，Unicode 字型呈現為「厩」，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 964 的「𪛗」字，Unicode 編碼為 28EF3，由於 Windows XP 並不支援，仍須使用構字式「𪛗𪛗焉」。

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、《新序》中未確定字形「●」共 7 處，詳細處理內容可參考附件二的未確定字形「●」手動取代表。附件二《新序》未確定字形「●」手動取代表欄位說明如下：

- 甲、檔案原文摘錄：摘錄檔案(新序.xml)中包含未確定字形「●」的文句段落，其中紅字底線的部分，為需要手動取代的字形。
- 乙、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。

附件一、《新序》造字分析表

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
826	癱	1	1	8E68	7655	癱		
964	隅	1	1	8F55	28EF3		𠃹𠃹焉	
972	犇	2	2	8F5D	7287	犇		
981	豺	1	1	8F66	72B2	豺		
1024	虵	8	8	8FB3	8675	虵		
1039	稟	2	2	8FC2	69C0	稟		
1067	碁	1	1	8FDE	68CA	碁		
1925	羈	1	1	9568	8989	羈		Unicode 呈現差異。
1981	讐	13	13	95C2	8B90	讐		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
1999	贊	1	1	95D4	8CDB	贊		
2152	蕓	1	1	96D0	979F	蕓		
2162	醎	1	1	96DA	918E	醎		
2415	鬪	4	4	987B	95D8	鬪		
2449		6	6	98BF			名𠂔隹	雜，校對問題，舊版漢籍錯字，可用程式全部取代。
2651	鸛	1	1	99EC	9E1C	鸛		
2812	舉	2	2	9AF0	64E7	舉		
2948	聵	3	3	9BDB	43C2		耳𠂔叟	
2969	譖	2	2	9BF0	8B5B	譖		
2997	𡗗	5	5	9C4D	25081		𠂔土人口皿𠂔	
3787	豕	1	1	8152	51A2	豕		
3789	冲	1	1	8154	51B2	冲		
3804	効	1	1	8163	52B9	効		
3813	邛	6	6	816C	536D	邛		
3840	坂	2	2	81A9	5742	坂		
3858	寃	1	1	81BB	5BC3	寃		
3868	卮	2	2	81C5	5DF5	卮		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
3883	徧	5	5	81D4	5FA7	徧		
3891	慙	3	3	81DC	6159	慙		
4012	鼓	1	1	82B8	76B7	鼓		
4038	竈	2	2	82D2	7AC8	竈		
4041	豎	3	3	82D5	7AEA	豎		
4055	粃	4	4	82E3	7C83	粃		
4070	縹	2	2	82F2	7DE4	縹		
4077	罇	1	1	82F9	7F47	罇		
4085	耻	1	1	8342	803B	耻		
4089	脇	2	2	8346	8107	脇		
4094	舘	1	1	834B	8218	舘		
4134	衛	2	2	8373	885E	衛		
4139	袴	1	1	8378	88B4	袴		
4144	霸	1	1	837D	8987	霸		Unicode 呈現差異。
4155	訶	1	1	83AA	8B0C	訶		
4156	諂	1	1	83AB	8B1F	諂		諂，校對問題，舊版漢籍錯字，可用程式全部取代。
4169	趲	1	1	83B8	8DA6	趲		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4184	迹	7	7	83C7	8FF9	迹		
4190	邳	4	4	83CD	90C4	邳		
4193	鈎	4	4	83D0	920E	鈎		
4237	髡	8	8	83FC	9AE0	髡		
4250	鰭	3	3	844A	9C0C	鰭		
4252	鰾	2	2	844C			魚𩺰眾	鰾，異體字問題。
4259	麇	1	1	8453	9E95	麇		Unicode 呈現差異。
4267	鼈	2	2	845B	9F08	鼈		
4283	戲	1	1	846B	622F	戲		
4309	刼	3	3	84A7	523C	刼		
4502	俎	2	2	85CB	723C	俎		
4512		1	1	85D5			口𠂔它	咤，校對問題，舊版漢籍錯字，可用程式全部取代。
4558	廐	3	3	8644	53A9	廐		Unicode 呈現差異。
4614	嫌	1	1	867C	3858		巾𠂔兼	
4768	妬	4	4	8779	59AC	妬		
4773	姉	1	1	877E	59C9	姉		Unicode 呈現差異。
4894		5	5	885A			宀𠂔心𠂔冉	寧，校對問題，舊版漢籍錯字，可用程式全部取代。

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4904	尅	1	1	8864	5C05	尅		
4907	橐	1	1	8867	69D6	橐		
5034	弃	1	1	8949	5F03	弃		Unicode 呈現差異。
5299	戟	1	1	8AD7	39B8		卓厶戈	戟，異體字問題。
5355	𡗗	2	2	8B50	20AE4		𡗗	
5439	栢	1	1	8BC6	6822	栢		
5541	蠹	1	1	8C6D	8827	蠹		
5591	杞	2	2	8CC1	233CC		木厶已	杞，異體字問題。
5702	厨	1	1	8D71	53A8	厨		

附件二、《新序》未確定字形「●」手動取代表

檔案原文摘錄	手動取代結果
先生老●歟	先生老 ¹ 厶昏歟
逐麋麋●鹿	逐麋麋 ^厶 鹿
爵也●擊之物	爵也 ^扌 厶瞿擊之物
良兔曰東郭●	良兔曰東郭 ^爰 厶兔
則雖東郭●	則雖東郭 ^爰 厶兔
兩手●地而歐之	兩手 ^扌 厶處地而歐之
●海內矣	^湓 海內矣