

《春秋繁露》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
2008/2/19 黃榮順 製作

一、《春秋繁露》一書(春秋繁露.xml)使用舊版造字 43 個，字頻 94 次，詳如附件一。這 43 個造字中，38 個可轉成 Windows XP 可顯示的字，字頻 88 次；另外 5 個字必須轉成構字式，字頻 6 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、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 5308 的「棹𣎵戈」字，原字應為「木𣎵戢」。
- 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 4134 的「衛」係「衛」字之異體，故以「衛」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 4145 的「羈」字，Unicode 字型呈現為「羈」，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。故如造字編號 963 的「鯪」字，Unicode 編碼為 4C78，Windows XP 並不支援，仍須改用構字式「魚𩺰叟」。

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、「春秋繁露.xml」發現之未確定字形「●」共 13 處，修正情形詳列附件二，附件二「春秋繁露.xml」「●」修改表體例說明如下：

甲、檔案原文摘錄：摘錄檔案(春秋繁露.xml)中包含「●」的文句段落，其中紅字底線的部分，為需要手動取代的原文。

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。

丙、備註：修改後的字形若屬於「待造字」，則於此欄標示。

附件一、「春秋繁露.xml」造字分析表

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
814	墓	5	5	8E5C	87C7	墓		
893	嵒	4	4	8ECD	5CE9	嵒		
907	嵒	1	1	8EDB	21F8B		山厶嵒	
910	嵒	1	1	8EDE	5D78	嵒		
963	鯨	1	1	8F54	4C78		魚厶叟	
982	胃	1	1	8F67	80F7	胃		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
1008	玦	1	1	8FA3	739E	玦		
1044	儻	7	7	8FC7	511B	儻		
1056	阤	1	1	8FD3	9641	阤		
1088	暘	4	4	8FF3	757C	暘		
1981	讐	1	1	95C2	8B90	讐		
1982	調	1	1	95C3	8B87	調		
2162	醎	2	2	96DA	918E	醎		
2478	鞏	1	1	98DC	978C	鞏		
2812	舉	2	2	9AF0	64E7	舉		
3806	勳	1	1	8165	52E0	勳		
3883	徧	10	10	81D4	5FA7	徧		
3910	擘	2	2	81EF	64E5	擘		
3961	灾	2	2	8263	707E	灾		
3965	裁	2	2	8267	70D6	裁		
3982	珣	1	1	8278	73F7	珣		
3996	疎	1	1	82A8	758E	疎		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4038	竈	1	1	82D2	7AC8	竈		
4040	竝	1	1	82D4	7ADD	竝		
4134	衛	1	1	8373	885E	衛		衛，異體字問題。
4140	衽	2	2	8379	88B5	衽		
4145	羈	1	1	837E	898A	羈		Unicode 字型差異。
4184	迹	2	2	83C7	8FF9	迹		
4237	髡	3	3	83FC	9AE0	髡		
4242	鬪	1	1	8442	9B2D	鬪		
4256	鷄	2	2	8450	9DC4	鷄		
4342	倮	4	4	84C8	502E	倮		
4371	倜	1	1	84E5	50F4	倜		
4558	厖	1	1	8644	53A9	厖		Unicode 字型差異。
4768	妬	5	5	8779	59AC	妬		
4981	皁	8	8	88D3	9CEC	皁		
5055	飪	1	1	895E	296C6		食 𪛗 𪛘	
5302	昏	1	1	8ADA	662C	昏		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
5308		2	2	8AE0			棹𠂔戈	木𠂔戢，舊漢籍錯字，可用程式全部取代。
5334	暉	1	1	8AFA	6692	暉		
5398	滄	2	2	8B7B	6E7B	滄		
5652	蛩	1	1	8CFE	27312		或𧈧虫	
5706	敍	1	1	8D75	654D	敍		敍，異體字問題。

附件二、「春秋繁露.xml」「●」修改表

檔案原文摘錄	手動取代結果	備註
大●者何	大𠂔者何	
大旱●祭而請	大旱𠂔祭而請	
園囿、萎●、臺沼	園囿、萎𠂔巷、臺沼	待造字
●眾惡於內	𠂔眾惡於內	
故心之為名，●也	故心之為名，𠂔也	
心何●哉	心何𠂔哉	
身有情欲●	身有情欲𠂔	
●。察實以為名	𠂔。察實以為名	
身首●而員	身首𠂔女而員	
絕源塞●	絕源塞𠂔	
剛毅，子●是也	剛毅，子𠂔是也	
摧嵬●巍	摧嵬𠂔罪巍	
尊實，●也	尊實，麥𠂔豐也	