

《鹽鐵論》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

2008/1/24 陳建安 製作

一、《鹽鐵論》一書(鹽鐵論.xml)使用舊版造字 69 個，字頻 152 次，詳如附件一。這 69 個造字中，51 個可轉成 Windows XP 能顯示的字，字頻 131 次；另外 18 個字必須轉成構字式，字頻 21 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「名𠂔隹」字，原字為「頤」。
- 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 5746 的「每𠂔系」係「鯨」字之異體，故以「鯨」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「廐」字，Unicode 字型呈現為「廐」，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。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 1689 的「禱」字，Unicode 碼為 464F，Windows XP 並不支援，故仍須使用構字式「禱 𠂔 專」。

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、《鹽鐵論》中未確定字形「●」共 17 處，詳細處理內容可參考附件二的未確定字形「●」手動取代表。附件二的未確定字形「●」手動取代表欄位說明如下：

- 甲、檔案原文摘錄：摘錄檔案(鹽鐵論.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、《鹽鐵論》造字分析表

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
827	𩺰	1	1	8E69	9B8C	𩺰		
845	𩺱	2	2	8E7B	9F26	𩺱		
881	𩺲	1	1	8EC1	98C7	𩺲		
899	𩺳	2	2	8ED3	99DE	𩺳		
1024	𩺴	2	2	8FB3	8675	𩺴		
1044	𩺵	2	2	8FC7	511B	𩺵		
1226	𩺶	4	4	90E0	7F78	𩺶		
1689	𩺷	1	1	93D8	464F		𩺷 𩺷	
1981	𩺸	1	1	95C2	8B90	𩺸		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
2160	磴	1	1	96D8	40D8		石 厶 堅	
2449		1	1	98BF			名 厶 隹	額，校對問題，舊版漢籍錯字，可用程式全部取代。
2524	饌	1	1	994B	994D	饌		
2659	𣎵	1	1	99F4	201D7		亻 厶 瓜	
3794	湊	1	1	8159	51D1	湊		
3798	𠂔	1	1	815D			𠂔	
3804	効	3	3	8163	52B9	効		
3840	坂	3	3	81A9	5742	坂		
3868	𡵓	1	1	81C5	5DF5	𡵓		
3883	徧	2	2	81D4	5FA7	徧		
3891	慙	1	1	81DC	6159	慙		
3894	筋	2	2	81DF	4225		𠂔 厶 勹	
3930	憵	2	2	8244	6901	憵		
3937	𣎵	2	2	824B	6A9D	𣎵		
3974	牀	3	3	8270	7240	牀		
3984	瑠	1	1	827A	7460	瑠		
4010	臯	2	2	82B6	7690	臯		Unicode 呈現差異。

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4014	蓋	2	2	82BA	8462	蓋		
4035	穽	3	3	82CF	7A7D	穽		
4077	罇	1	1	82F9	7F47	罇		
4085	耻	7	7	8342	803B	耻		
4091	皐	4	4	8348	81EF	皐		Unicode 呈現差異。
4095	舩	1	1	834C	8229	舩		
4134	衛	10	10	8373	885E	衛		
4139	袴	2	2	8378	88B4	袴		
4140	衽	1	1	8379	88B5	衽		
4146	覩	25	25	83A1	89A9	覩		
4150	訕	2	2	83A5	8A29	訕		
4172	屨	1	1	83BB	8E77	屨		Unicode 呈現差異。
4184	迹	3	3	83C7	8FF9	迹		
4222	顙	1	1	83ED	9847	顙		
4226	飡	2	2	83F1	98E1	飡		
4234	羸	2	2	83F9	9A58	羸		
4242	鬪	3	3	8442	9B2D	鬪		Unicode 呈現差異。
4265	鼃	3	3	8459	9F03	鼃		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4267	鼈	2	2	845B	9F08	鼈		
4458	勸	2	2	857D	52CC	勸		Unicode 呈現差異。
4558	厖	1	1	8644	53A9	厖		Unicode 呈現差異。
4561		1	1	8647			𠂇 厶 曲 比 𠂇	厖，校對問題，舊版漢籍錯字，可用程式全部取代。
4727	埤	1	1	8750	58C4	埤		
4768	妬	5	5	8779	59AC	妬		
4825	闢	1	1	87D4	28D66		門 厶 俞	
4887	醯	1	1	8853	4908		酉 厶 盆	
4910		1	1	886A			氵 厶 𩚑	淫，校對問題，舊版漢籍錯字，可用程式全部取代。
4981	鳧	1	1	88D3	9CEC	鳧		鳧，校對問題，舊版漢籍錯字，可用程式全部取代。
4996	斲	1	1	88E2	65B5	斲		
5045	飪	1	1	8954	9901	飪		
5087	躐	1	1	897E	2810C		足 厶 習	
5265	蕎	1	1	8AB5	26E1C		艹 厶 𠂇 𠂇 回	
5308		2	2	8AE0			棹 厶 戈	檝，校對問題，舊版漢籍錯字，可用程式全部取代。
5434	湟	1	1	8BC1	5E7C		形 氵 白 工 𠂇	

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
5466	倏	1	1	8BE1	5010	倏		Unicode 呈現差異。
5497	衤	1	1	8C41	8873	衤		
5588	忤	1	1	8CBE	6061	忤		
5722		2	2	8DA7				𦵒，校對問題，舊版漢籍錯字，可用程式全部取代。
5723	𦵒	1	1	8DA8	4367		𦵒 𦵒 羊	𦵒，校對問題，舊版漢籍錯字，可用程式全部取代。
5724	𦵒	2	2	8DA9	7A38	𦵒		
5746		1	1	8DBF			𦵒 𦵒 系	𦵒，異體字問題。
5747	𦵒	1	1	8DC0	3FCA		𦵒 𦵒 晉	
5764	𦵒	2	2	8DD1	7C5E	𦵒		

附件二、《鹽鐵論》未確定字形「●」手動取代表

檔案原文摘錄	手動取代結果
樹木數徙則●	樹木數徙則𦵒 𦵒 委
鮮羔●，幾胎肩	鮮羔𦵒 𦵒 𦵒，幾胎肩
豐●耳菜	豐𦵒 𦵒 𦵒耳菜
𦵒●壁飾	𦵒 𦵒 𦵒壁飾
大夫士則單●木具	大夫士則單𦵒木具
燕●代黃	燕𦵒 𦵒 𦵒代黃
富者●耳銀鑷●	富者𦵒 𦵒 𦵒耳銀鑷𦵒 𦵒 𦵒
華●明鮮	華𦵒 𦵒 𦵒明鮮
中者●皮代旃	中者𦵒 𦵒 𦵒皮代旃
狗●馬腹	狗𦵒 𦵒 𦵒馬腹
糠●之苦也	糠𦵒 𦵒 𦵒之苦也

檔案原文摘錄	手動取代結果
不知老母之●賴	不知老母之顯賴
犀●咒甲	犀革△由咒甲
以道德為●	以道德為革△由
●裂可以補	巾△祭裂可以補
切而不●	切而不火△索