



《風俗通義》舊版造字轉碼說明

中研院資訊所文獻處理實驗室
中央研究院語言所文獻語料小組
2008/12/11 丁致伶

轉碼主要工作是把檔案中的舊版造字轉換成 Windows XP 能支援的 Unicode 字形，Unicode 目前共收錄漢字 70194 個字，而 XP 只能支援 20902 個字（詳如表一），不支援之字將以構字式表達。例：造字編號 960 的「聿」字，Unicode 編碼是 26553，由於 XP 並不支援，仍需使用構字式「工几耳」。

表一、Unicode 的字數及編碼區段

Unicode	新增字數	新增編碼區段	總字數	WinXP
1.1 版	20902	4E00-9FFF	20902	支援
3.0 版	6582	3400-4DFF	27484	不支援
3.1 版	42710	20000-2A6D6	70194	不支援

一、舊版造字轉碼分析：

《風俗通義》使用舊版造字 75 個，字頻 213 次，這 75 個造字中，64 個可轉成 Windows XP 能顯示的字，字頻 200 次；另外 11 個字必須轉成構字式，字頻 13 次。

轉碼完成製作轉碼分析表，請參考附件一《風俗通義》轉碼分析表，欄位說明如下：

- (一) 編號：Big5 造字空間為 6217 個，編號由 1 到 6217。
- (二) 造字：舊版造字。
- (三) 頻次：舊版造字在文件的出現次數。
- (四) Big5：造字的 Big5 碼。
- (五) Unicode：造字所對應的 Unicode 碼。

(六) WinXP：造字在 Windows XP 的對應字形。

(七) 構字式：Windows XP 無法對應字形改用構字式。

(八) 備註凡例：備註欄中記錄轉碼後字形及修改原因，凡例如下：

1. 異體字問題：為了使用者查詢和使用的方便，在處理異體字時最主要的方式是以標準字取代，除非是專有名詞或特殊情形，如：人名、地名等。例：造字編號 4134 的「衛」字，是「衛」的異體字，以標準字「衛」取代。
2. 錯字，以程式全部取代：與原書字形不符，並查詢教育部異體字字典確認非異體字後，皆歸類為錯字，以程式取代為正確字形。例：檔案中使用造字編號 4180 的「臯」字，而原書使用之字為「臯」，查詢教育部異體字字典確認「臯」並非「臯」之異體字，因此歸為錯字，以程式取代為正確字形「臯」。
3. Unicode 字型呈現差異：Unicode 字型與舊版造字有些微差異，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 3935 的「槩」字，Unicode 字型呈現為「槩」，實際上仍為同一字。

附件一、《風俗通義》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
826	癱	1	8E68	7655	癱		
832	沱	2	8E6E	6CB2	沱		
852	貉	1	8EA4	72E2	貉		
960	聾	1	8F51	26553		形工几 耳	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
964	馮	1	8F55	28EF3		卩厶焉	
977	磨	3	8F62	9E8F	磨		
994	楚	1	8F73	8E75	楚		
1024	虵	12	8FB3	8675	虵		
1044	儻	2	8FC7	511B	儻		
1097	嶠	1	8FFC	37ED		山厶昏	
1226	罰	5	90E0	7F78	罰		
1642	脉	1	93A9	8109	脉		
1650	賭	1	93B1	4417		月厶者	
1893	袞	1	9548	886E	袞		
1981	讐	3	95C2	8B90	讐		讐，異體字問題。
3133	癩	1	9CF7	7651	癩		
3787	冢	7	8152	51A2	冢		
3804	効	1	8163	52B9	効		
3814	却	2	816D	5374	却		
3835	噍	1	81A4	564D	噍		
3840	坂	1	81A9	5742	坂		
3849	獎	1	81B2	596C	獎		獎，異體字問

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							題。
3858	冤	2	81BB	5BC3	冤		冤，異體字問題。
3883	徧	6	81D4	5FA7	徧		
3891	慙	4	81DC	6159	慙		
3921	皙	1	81FA	6673	皙		
3935	槩	1	8249	3BA3	槩		Unicode 字型呈現差異。
3974	牀	4	8270	7240	牀		
3991	攜	1	82A3	64D5	攜		
4016	着	1	82BC	7740	着		
4028	稟	1	82C8	7980	稟		
4035	穽	2	82CF	7A7D	穽		
4038	竈	10	82D2	7AC8	竈		
4041	豎	1	82D5	7AEA	豎		
4080	羣	4	82FC	7FA3	羣		
4085	耻	2	8342	803B	耻		
4090	脚	1	8347	811A	脚		
4091	阜	2	8348	81EF	阜		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4113	簡	1	835E			𠄎 𠄎 間	
4131	𠄎	1	8370	8842	𠄎		
4132		24	8371			血 𠄎 𠄎	眾，異體字問題。
4134	衛	11	8373	885E	衛		衛，異體字問題。
4180	臯	1	83C3	8FA0	臯		臯，錯字，以程式全部取代。
4184	迹	3	83C7	8FF9	迹		
4237	髡	5	83FC	9AE0	髡		
4242	鬪	5	8442	9B2D	鬪		
4256	鷄	15	8450	9DC4	鷄		雞，異體字問題。
4309	刼	1	84A7	523C	刼		
4455	劒	1	857A			僉 𠄎 刀	
4460	勅	9	85A1	52D1	勅		
4507	吳	10	85D0	5434	吳		吳，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4716	堙	1	8745	57FF	堙		
4723	塲	1	874C	58C3	塲		
4762	旃	2	8773	3ACB		扌△冉	
4858	閱	1	87F5	49A7		門△兒	閱，異體字問題。
4904	尅	3	8864	5C05	尅		
4996	斲	1	88E2	65B5	斲		
5019	畧	1	88F9			畧	
5034	弃	1	8949	5F03	弃		棄，異體字問題。
5085	苒	1	897C	26B12		艹△丰	半，異體字問題。
5131	窻	1	89CC	7ABB	窻		
5141	昊	1	89D6	23154		昊	
5225	邨	2	8A6B	460F		血△卩	
5289	慤	1	8ACD	6164	慤		
5310	烏	1	8AE2	8204	烏		
5319	晰	1	8AEB	6663	晰		
5354	昊	1	8B4F	23173		日△火	昊，異體字問題

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							題。
5439	栢	3	8BC6	6822	栢		
5452	轆	1	8BD3	8F5C	轆		
5482	璽	1	8BF1	8812	璽		
5559	欸	1	8CA1	6B35	欸		
5588	恠	2	8CBE	6061	恠		
5702	厨	1	8D71	53A8	厨		
5706	敘	3	8D75	654D	敘		敘，異體字問題。
5742	澁	1	8DBB	6F81	澁		

二、手動取代表說明：

《風俗通義》檔案中出現「●」字形，共計17字，比對原書後以Unicode字形或構字式手動取代，並製作「●」字手動取代表，請參考附件二。欄位說明如下：

1. 頁碼：「●」字形所在位置之原書頁碼。
2. 檔案原文摘錄：摘錄《風俗通義》檔案有「●」字形的文句。
3. 原書字形：「●」在原書中之字形，以此字取代檔案原文的「●」。
4. 備註：記錄修改相同詞句的次數或其他事項。

附件二、《風俗通義》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 3	悉地力，種●	殷△木	
p. 162	負●●涉	𠂇△肉 山△升	
p. 196	●宅而居之	卩△聶	
p. 228	樂安周●孟玉	王△卩	
p. 228	●亦自劾去	王△卩	
p. 230	周●苟執果毅	王△卩	
p. 281	大笙謂之●	𠂇△巢	
p. 313	●	簌	有 2 處
p. 313	●，簫也	簌	
p. 313	言其聲音●●	簌簌	
p. 364	以●燎祀風師	示△亞	待造字
p. 367	過於●上	形：𠂇 𠂇 回	
p. 378	食新曰●臚	臚	
P. 441	往手●摸之	扌△攴	