

《白虎通疏證》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
2008/12/04 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 4825 的「闢」字，Unicode 編碼是 28D66，由於 XP 並不支援，仍需使用構字式「門△俞」。

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：

《白虎通疏證》使用舊版造字 53 個，字頻 166 次，這 53 個造字中，50 個可轉成 Windows XP 能顯示的字，字頻 160 次；另外 3 個字必須轉成構字式，字頻 6 次。

轉碼完成製作轉碼分析表，請參考附件一《白虎通疏證》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 3796 的「決」字，是「決」的異體字，以標準字「決」取代。
2. 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 4052 的「簾」字，Unicode 字型呈現為「簾」，實際上仍為同一字。

附件一、《白虎通疏證》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
1044	儻	4	8FC7	511B	儻		
1226	罰	7	90E0	7F78	罰		
1530	紉	2	92D6	7CFA	紉		
1578		1	9347			日 𠄎 冉	𠄎，異體字問題。
1642	脉	1	93A9	8109	脉		
1893	袞	2	9548	886E	袞		
1906	袞	5	9555	88E6	袞		
2140	迫	2	96C4	5EF9	迫		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
2163	杼	1	96DB	7CB0	杼		
2811		3	9AEF			𠄎 𠄎 田	畱，異體字問題。
2863		2	9B64			𠄎 𠄎 友	茈，異體字問題。
3255	善	18	9DD4	5584	善		
3787	冢	1	8152	51A2	冢		
3791	况	2	8156	51B5	况		
3796	決	2	815B	51B3	決		決，異體字問題。
3804	効	2	8163	52B9	効		
3806	勳	2	8165	52E0	勳		
3826	叶	2	8179	53F6	叶		
3846	壻	10	81AF	58FB	壻		
3883	徧	7	81D4	5FA7	徧		
3924	菴	1	81FD	671E	菴		
3930	桴	2	8244	6901	桴		
3974	牀	1	8270	7240	牀		
3996	疎	5	82A8	758E	疎		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4010	皐	2	82B6	7690	皐		皐，異體字問題。
4011	皞	4	82B7	76A1	皞		皞，異體字問題。
4038	竈	7	82D2	7AC8	竈		
4043	筭	1	82D7	7B53	筭		
4052	簾	1	82E0	7C34	簾		Unicode 字型呈現差異。
4073	纒	1	82F5	7E6E	纒		
4080	羣	16	82FC	7FA3	羣		
4091	皐	3	8348	81EF	皐		
4107	葱	4	8358	8471	葱		
4132		5	8371			血𠂇𠂇	眾，異體字問題。
4134	衛	6	8373	885E	衛		衛，異體字問題。
4140	衽	1	8379	88B5	衽		
4142	裛	1	837B	8943	裛		
4146	覩	1	83A1	89A9	覩		睹，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							題。
4178	輓	1	83C1	8F2D	輓		
4184	迹	3	83C7	8FF9	迹		
4242	鬪	2	8442	9B2D	鬪		
4469	勻	1	85AA	5304	勻		
4507	吳	5	85D0	5434	吳		吳，異體字問題。
4768	妬	2	8779	59AC	妬		妒，異體字問題。
4825	鬪	2	87D4	28D66		門△俞	
5019	畧	3	88F9			畧	
5173	漱	1	89F6	6F44	漱		
5310	烏	1	8AE2	8204	烏		
5355	𡗗	1	8B50	20AE4		𡗗	
5591	杞	2	8CC1	233CC		木△巳	杞，異體字問題。
5702	厨	1	8D71	53A8	厨		
5706	敍	2	8D75	654D	敍		敘，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5860	无	1	C6D3	65E0	无		

二、手動取代表說明：

《白虎通疏證》檔案中出現「●」字形，共計29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《白虎通疏證》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、《白虎通疏證》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 50	臥之●●	詒詒	
p. 54	謂之堯者何？堯猶●	堯	
p. 54	●也。至高之貌。	堯	
p. 54	舜猶●●也	僣僣	
p. 111	恐有過●也	言𠄎吳	改為誤
p. 237	民●毒螫	业△冢	改為蒙
p. 255	匪我求童●	业△冢	改為蒙
p. 255	童●求我	业△冢	改為蒙
p. 284	阜出●莆	形 ^{tt} トヨル□	改為萑

頁碼	檔案原文摘錄	原書字形	備註
p. 285	則●蕭生庖厨	形 ^艹 卜ヨ匕 ^口	改為蕙
p. 285	●蕭者	形 ^艹 卜ヨ匕 ^口	改為蕙
p. 342	昌●風至	盍	
p. 342	昌●者	盍	
p. 343	昌●風至	盍	有 2 處
p. 351	地道安●而出財物	寧	
p. 388	青目●張	目 ^厶 辰	
p. 392	天用●絕其命矣	巢 ^厶 刀	改為剿
p. 393	儻儻然如●家之狗	奘 ^厶 以	改為喪
p. 393	如●家之狗	奘 ^厶 以	改為喪
p. 431	三年之●	奘 ^厶 以	改為喪
p. 436	去●無所不佩	奘 ^厶 以	改為喪
p. 494	天子大夫赤●葱衡	糸 ^厶 友	改為紱
p. 495	冠者，●也	巾 ^厶 卷	
p. 495	所以●持其髮者也	巾 ^厶 卷	
p. 505	三年之●達乎天子	奘 ^厶 以	改為喪
p. 534	●然伏僵	卜 ^厶 崩	待造字