

《宜春香質》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
2008/7/15 丁玟伶

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 2694 的「餽」字，Unicode 編碼是 296DE，由於 XP 並不支援，仍需使用構字式「食 𠂔 孚」。

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：

《宜春香質》使用舊版造字 56 個，字頻 547 次，這 56 個造字中，47 個可轉成 Windows XP 能顯示的字，字頻 463 次；另外 9 個字必須轉成構字式，字頻 84 次。

轉碼完成製作轉碼分析表，請參考附件一《宜春香質》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 3791 的「況」字，是「況」的異體字，以標準字「況」取代。

附件一、《宜春香質》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
296	厶	9	FBEC			厶	
297	厶	13	FBED			厶	
298	厶	50	FBEE			厶	
658	彡	1	FE5D			彡	
660	彡	9	FE5F	4E02	彡		
766	彡	2	FEEB			彡	
770	彡	3	FEEF			彡	
780	彡	1	FEF9			彡	
782	彡	1	FEFB	5E7A	彡		
2046	踪	1	9644	8E2A	踪		
2053	躡	1	964B	8EA7	躡		
2346	鐳	1	97F5	9427	鐳		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
2694	餗	1	9A58	296DE		食 𠂇 孚	
3787	冢	1	8152	51A2	冢		
3789	冲	5	8154	51B2	冲		沖，異體字問題。
3791	况	9	8156	51B5	况		況，異體字問題。
3805	勅	1	8164	52C5	勅		
3815	𡗗	1	816E	537A	𡗗		
3824	叙	4	8177	53D9	叙		
3825	疊	3	8178	53E0	疊		疊，異體字問題。
3829	咏	2	817C	548F	咏		
3830	咤	1	817D	54A4	咤		
3868	卮	2	81C5	5DF5	卮		
3886	悞	1	81D7	609E	悞		
3930	憵	1	8244	6901	憵		
3938	橈	5	824C	6AC8	橈		
3966	烟	11	8268	70DF	烟		
3968	煨	1	826A	7145	煨		
3974	牀	54	8270	7240	牀		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3996	疎	2	82A8	758E	疎		
4000	疴	1	82AC	75B4	疴		
4015	眦	1	82BB	7726	眦		
4016	着	190	82BC	7740	着		
4039	疇	1	82D3	7ADA	疇		
4065	紮	1	82ED	7D25	紮		
4067	綉	1	82EF	7D89	綉		
4068	綫	5	82F0	7DAB	綫		
4073	纒	1	82F5	7E6E	纒		
4074	縑	2	82F6	7E27	縑		
4085	耻	4	8342	803B	耻		
4089	脇	1	8346	8107	脇		
4129	虬	1	836E	866C	虬		
4132		50	8371			血𠂇𠂇	眾，異體字問題。
4139	袴	14	8378	88B4	袴		
4167	賈	1	83B6	8CEB	賈		
4173	駝	1	83BC	8EAD	駝		
4193	鈎	5	83D0	920E	鈎		鈎，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4195	鉢	2	83D2	9262	鉢		
4232	隲	1	83F7	96B2	隲		
4256	鷄	10	8450	9DC4	鷄		雞，異體字問題。
4307	罵	12	84A5	99E1	罵		罵，異體字問題。
4518	响	1	85DB	54CD	响		
4768	妬	5	8779	59AC	妬		妒，異體字問題。
4801	斌	2	87BC	5A2C	斌		
5075	彡	4	8972			彡	
5621	泪	33	8CDF	6CEA	泪		淚，異體字問題。