

《隋唐嘉話》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
2008/4/10 陳建安 製作

一、《隋唐嘉話》一書(隋唐嘉話.xml)使用舊版造字 43 個，字頻 89 次，詳如附件一。這 43 個造字中，35 個可轉成 Windows XP 能顯示的字，字頻 73 次；另外 8 個字必須轉成構字式，字頻 16 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、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 3800 的「杀𠄎」係「剎」字之異體，故以「剎」字取代。又編號 3966 的「烟」字，係「煙」字之簡體字，亦以標準字的「煙」字取代。

2、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 1076 的「高」字，Unicode 字型呈現為「高」，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。所以造字編號 2302 的「忸」字，Unicode 編碼為 38FF，由於 Windows XP 並不支援，仍須使用構字式「忸𠄎勺」。

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

附件一、《隋唐嘉話》造字分析表

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
297	𠂔	1	1	FBED			𠂔	
298	𠂕	1	1	FBEE			𠂕	
759	广	1	1	FEE4			广	
1076	高	1	1	8FE7	5368	高		Unicode 呈現差異。
1226	罰	2	2	90E0	7F78	罰		
1530	紕	1	1	92D6	7CFA	紕		
1642	脉	2	2	93A9	8109	脉		
1981	讐	1	1	95C2	8B90	讐		
2056	佺	2	2	964E	4FD2	佺		
2302	忉	1	1	97C9	38FF		忉 𠂔 勺	
2651	鸛	1	1	99EC	9E1C	鸛		
3020	衤	1	1	9C64			衤	
3780	僑	1	1	814B	5101	僑		
3789	冲	4	4	8154	51B2	冲		冲，異體字問題。
3800		1	1	815F			杀 𠂔 刂 刹	刹，異體字問題。

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
3801	剗	1	1	8160	5257	剗		
3804	効	1	1	8163	52B9	効		
3814	却	4	4	816D	5374	却		
3876	廻	1	1	81CD	5EFB	廻		迴，異體字問題。
3931	椀	1	1	8245	6900	椀		
3935	槩	1	1	8249	69E9	槩		概，異體字問題。
3966	烟	1	1	8268	70DF	烟		煙，異體字問題。
3974	牀	1	1	8270	7240	牀		
3990	甄	1	1	82A2	750E	甄		
3996	疎	1	1	82A8	758E	疎		
4016	着	3	3	82BC	7740	着		著，異體字問題。
4080	羣	14	14	82FC	7FA3	羣		群，異體字問題。
4090	脚	1	1	8347	811A	脚		
4107	葱	1	1	8358	8471	葱		
4115	藁	1	1	8360	85C1	藁		
4129	虬	2	2	836E	866C	虬		
4132		7	7	8371			血𠂇𠂇眾	眾，異體字問題。
4167	賁	1	1	83B6	8CEB	賁		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4184	迹	3	3	83C7	8FF9	迹		
4242	鬪	1	1	8442	9B2D	鬪		Unicode 呈現差異。
4460	勅	1	1	85A1	52D1	勅		
4507	吳	1	1	85D0	5434	吳		吳，異體字問題。
4530	咬	2	2	85E7	6648	咬		
4768	妬	5	5	8779	59AC	妬		妒，異體字問題。
4996	斲	1	1	88E2	65B5	斲		
5176		3	3	89F9			汜 𠄎 冗	沉，異體字問題。
5705	廐	1	1	8D74	5ED0	廐		廐，異體字問題。
5706	敍	7	7	8D75	654D	敍		敍，異體字問題。