

《王梵志詩》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
2008/3/20 黃榮順 製作

一、《王梵志詩》一書(王梵志詩.xml)使用舊版造字 65 個，字頻 167 次，詳如附件一。這 65 個造字中，51 個可轉成 Windows XP 可顯示的字，字頻 130 次；另外 14 個字必須轉成構字式，字頻 37 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、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 2834 的「攵」字，原字應為「攴」。
- 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 3858 的「冤」係「冤」字之異體，故以「冤」字取代。又如編號 5034 的「弃」係「棄」字之簡體，故以「棄」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 1536 的「𪚩」字，Unicode 字型呈現為「𪚩」，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。所以如造字編號 1598 的「𪚩」字，

Unicode 編碼為 4A7A，由於 Windows XP 並不支援，仍須改用構字式「革𠂔翁」。

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、「王梵志詩.xml」造字分析表

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
296	𠂔	3	3	FBEC			𠂔	
297	𠂔	13	13	FBED			𠂔	
298	𠂔	2	2	FBEE			𠂔	
299	𠂔	2	2	FBEF			𠂔	
300	𠂔	2	2	FBF0			𠂔	
717	𠂔	2	2	FEBA			𠂔	
829	𠂔	4	4	8E6B	81F0	𠂔		
842	𠂔	6	6	8E78	3AB7		𠂔	
1014	𠂔	1	1	8FA9	73CE	𠂔		
1039	𠂔	1	1	8FC2	69C0	𠂔		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
1056	阨	1	1	8FD3	9641	阨		
1081	鶴	1	1	8FEC	9E16	鶴		
1536	麤	3	3	92DC	9E81	麤		Unicode 字型差異。
1557	羴	1	1	92F1			羊𠂔歷	
1598	鞫	1	1	935B	4A7A		革𠂔翁	
1772		1	1	946C			𠂔𠂔梨	藜，異體字問題。
1800	菌	1	1	94AA	8597	菌		
2021	趲	3	3	95EA	8D82	趲		
2044	踈	5	5	9642	8E08	踈		
2408	鬧	1	1	9874	9599	鬧		
2430	𧇧	1	1	98AC	473A		豆𠂔昔	
2467	壿	1	1	98D1	367A		土𠂔匱	
2631	鷄	1	1	99D8	9D44	鷄		
2834	攵	2	2	9B47	6535	攵		攵，校對問題，舊版漢籍錯字，可用程式全部取代。
3803	𩚑	1	1	8162	4B54		食𠂔追	
3858	寃	1	1	81BB	5BC3	寃		冤，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
3891	慙	3	3	81DC	6159	慙		
3897	慙	3	3	81E2	615C	慙		
3931	椀	1	1	8245	6900	椀		
3935	槩	3	3	8249	69E9	槩		概，異體字問題。
3961	灾	3	3	8263	707E	灾		災，異體字問題。
3974	牀	3	3	8270	7240	牀		
4038	竈	2	2	82D2	7AC8	竈		
4041	豎	1	1	82D5	7AEA	豎		
4044	筋	1	1	82D8	7B6F	筋		
4059	糢	1	1	82E7	7CC7	糢		
4077	罇	1	1	82F9	7F47	罇		
4094	舘	1	1	834B	8218	舘		館，異體字問題。
4105	菓	1	1	8356	83D3	菓		
4109	葛	1	1	835A	84AD	葛		
4118	藥	1	1	8363	8602	藥		
4139	袴	1	1	8378	88B4	袴		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4146	覩	2	2	83A1	89A9	覩		
4173	𪗇	1	1	83BC	8EAD	𪗇		
4195	鉢	3	3	83D2	9262	鉢		
4226	𪗇	9	9	83F1	98E1	𪗇		
4242	鬪	5	5	8442	9B2D	鬪		
4345		1	1	84CB			口𪗇𪗇	嗽，異體字問題。
4546	鷗	1	1	85F7	9D76	鷗		
4716	埜	1	1	8745	57FF	埜		
5008	幪	1	1	88EE	5E5E	幪		
5034	弃	5	5	8949	5F03	弃		棄，異體字問題。
5071	忝	4	4	896E	5FE9	忝		
5099	惣	26	26	89AC	60E3	惣		
5105	埳	1	1	89B2	5860	埳		
5133	憇	1	1	89CE	6199	憇		
5147	恠	3	3	89DC	6060	恠		
5179	駟	7	7	89FC	99C8	駟		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
5339	𪗇	1	1	8B40	59E7	𪗇		
5450	𪗈	1	1	8BD1	6E4C	𪗈		
5542	𪗉	1	1	8C6E	93C1	𪗉		
5552	𪗊	1	1	8C78	5F06	𪗊		
5555	迥	2	2	8C7B	9008	迥		
5588	𪗋	1	1	8CBE	6061	𪗋		
5704	𪗌	1	1	8D73	6C0A	𪗌		