

《西京雜記》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
2008/6/16 陳維君 製作

一、《西京雜記》一書(西京雜記.xml)使用舊版造字 79 個，字頻 159 次，詳如附件一。這 79 個造字中，64 個可轉成 Windows XP 能顯示的字，字頻 134 次；另外 15 個字必須轉成構字式，字頻 25 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 1067 的「碁」係「棋」字之異體，故以「棋」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 4242 的「鬪」字，Unicode 字型呈現為「鬪」，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 4101 的「珣」字，Unicode 編碼為 3EC1，由於 Windows XP 並不支援，仍須使用構字式「王𠂔回」。

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

四、《西京雜記》一書(西京雜記.xml)中未確定字形「●」共出現 9 次，詳細取代內容可參考附件二未確定字形「●」手動取代表。附件二未確定字形「●」手動取代表的欄位說明如下：

- 甲、 頁碼：「●」字形所在位置之原書頁碼。
- 乙、 檔案原文摘錄：摘錄檔案(西京雜記.xml)中包含該造字的文句段落，其中紅字底線的部分，為需要手動取代的原文。
- 丙、 手動取代結果：變更過後的檔案內容，其中藍字底線的部分，為修改後的結果。
- 丁、 備註：修改後的字形若屬於「待造字」，則於此欄標示。目前資料庫無法處理之字形仍維持「●」。

五、對於在缺字之外發現的錯字，詳細取代內容可參考附件三校出錯字表。附件三校出錯字表的欄位說明說明如下：

- 甲、 檔案原文摘錄：摘錄檔案(西京雜記.xml)中包含錯字的文句段落，其中紅字底線的部分，為需要手動取代的字形。
- 乙、 手動取代結果：變更過後的檔案內容，其中藍字底線的部分，為修改後的結果。

附件一、《西京雜記》造字分析表

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
842	𣪗	2	2	8E78	3AB7		𣪗	
1056	阨	1	1	8FD3	9641	阨		
1067	碁	1	1	8FDE	68CA	碁		碁，異體字問題。
1238	侯	2	2	90EC	77E6	侯		侯，異體字問題。
1266	碓	1	1	9149	784B	碓		
1664	腸	1	1	93BF	8193	腸		腸，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
1999	贊	1	1	95D4	8CDB	贊		贊，異體字問題。
2044	踈	1	1	9642	8E08	踈		
2278	鋤	1	1	97B1	936E	鋤		
2325	潤	2	2	97E0	77A4	潤		
2415	鬪	1	1	987B	95D8	鬪		鬪，異體字問題。
2516	餽	1	1	9943	9919	餽		
2650	鸛	1	1	99EB	9E18	鸛		
3035	賔	5	5	9C73	8CD4	賔		賔，異體字問題。
3787	冢	3	3	8152	51A2	冢		
3791	况	1	1	8156	51B5	况		况，異體字問題。
3796	決	2	2	815B	51B3	決		決，異體字問題。
3805	勅	1	1	8164	52C5	勅		
3817	廁	3	3	8170	53A0	廁		廁，異體字問題。
3840	坂	1	1	81A9	5742	坂		
3868	卮	1	1	81C5	5DF5	卮		
3876	廻	6	6	81CD	5EFB	廻		廻，異體字問題。
3877	迺	1	1	81CE	5EFC	迺		迺，異體字問題。
3883	徧	1	1	81D4	5FA7	徧		

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		慙，異體字問題。
3974	牀	14	14	8270	7240	牀		
4012	鼓	3	3	82B8	76B7	鼓		
4013	盃	1	1	82B9	76CC	盃		
4014	蓋	7	7	82BA	8462	蓋		蓋，異體字問題。
4057	墓	7	7	82E5	585F	墓		
4071	纓	2	2	82F3	7DF5	纓		
4077	罇	1	1	82F9	7F47	罇		
4080	羣	6	6	82FC	7FA3	羣		羣，異體字問題。
4085	耻	1	1	8342	803B	耻		
4089	脇	1	1	8346	8107	脇		脅，異體字問題。
4090	脚	4	4	8347	811A	脚		腳，異體字問題。
4091	臯	3	3	8348	81EF	臯		
4101	珣	1	1	8352	3EC1		王𠂔回	
4129	虬	1	1	836E	866C	虬		
4132		2	2	8371			血𠂔𠂔	眾，異體字問題。
4180	臯	1	1	83C3	8FA0	臯		
4184	迹	2	2	83C7	8FF9	迹		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
4193	鈎	1	1	83D0	920E	鈎		鈎，異體字問題。
4203	隣	2	2	83DA	96A3	隣		鄰，異體字問題。
4240	鬚	1	1	8440	9AF4	鬚		
4242	鬪	2	2	8442	9B2D	鬪		Unicode 呈現差異。
4254	鷓	1	1	844E	9D7E	鷓		
4256	鷄	3	3	8450	9DC4	鷄		
4267	鼈	1	1	845B	9F08	鼈		
4278	亘	1	1	8466	4E98	亘		
4297	躔	1	1	8479	281E0		形足厂墨	躔，異體字問題。
4345		1	1	84CB			口 ㄩ 敕	嗽，異體字問題。
4389	瑫	2	2	84F7	746B	瑫		
4455	劒	6	6	857A			僉 ㄩ 刀	
4496	段	1	1	85C5	53DA	段		段，異體字問題。
4502	俎	1	1	85CB	723C	俎		
4515	呪	1	1	85D8	546A	呪		
4531	攢	1	1	85E8	6505	攢		攢，異體字問題。
4552	趣	1	1	85FD	27F08		走 ㄩ 取	趣，異體字問題。
4607	狫	1	1	8675	3E60		犴 ㄩ 屯	

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
4643	徃	7	7	86BB	5F83	徃		徃，異體字問題。
4704	犝	1	1	86F8	7271	犝		
4770	鑄	2	2	877B	28A8B		金 𠂔 冫	
4812	羴	2	2	87C7	26351		羴 𠂔、	羴，異體字問題。
4907	橐	1	1	8867	69D6	橐		橐，異體字問題。
5072	扃	1	1	896F	6244	扃		扃，異體字問題。
5096	紆	2	2	89A9	42B8		纟 𠂔 亏	紆，異體字問題。
5131	窻	1	1	89CC	7ABB	窻		Unicode 呈現差異。
5176		1	1	89F9			冫 𠂔 冗	沉，異體字問題。
5290	携	1	1	8ACE	643A	携		
5310	烏	1	1	8AE2	8204	烏		
5338	暎	2	2	8AFE	668E	暎		
5439	栢	1	1	8BC6	6822	栢		
5533		1	1	8C65			殸 𠂔 米	穀，異體字問題。
5559	欸	2	2	8CA1	6B35	欸		欸，異體字問題。
5702	厨	1	1	8D71	53A8	厨		廚，異體字問題。
5719	麈	1	1	8DA4	4D25		鹿 𠂔 加	
5723	牂	1	1	8DA8	4367		牛 𠂔 羊	

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
5764	籩	1	1	8DD1	7C5E	籩		

附件二、《西京雜記》「●」字形手動取代表

頁碼	檔案原文摘錄	手動取代結果	備註
1074-1	白俞●杜●桂蜀	白俞木△匊杜木△匊桂蜀	待造字
1078-1	●徒驪山	送徒驪山	
1083-2	奮皓翅之●●	奮皓翅之羽△戈羽△戈	
1084-1	載●載齊	載△曾載齊	
1086-1	亦善為歸風●送遠之操	亦善為歸風送遠之操	●為多餘符號，人工刪除。
1086-1	倍升為●。倍●為紀	倍升為△式。倍△式為紀	

附件三、《西京雜記》校出錯字表

檔案原文摘錄	手動取代結果
則見腸胃五臟	則見腸胃五臟