

《新校搜神記》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
2008/6/12 黃榮順 製作

一、《新校搜神記》一書(新校搜神記.xml)使用舊版造字 83 個，字頻 305 次，詳如附件一。這 83 個造字中，67 個可轉成 Windows XP 可顯示的字，字頻 255 次；另外 16 個字必須轉成構字式，字頻 50 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 3819 的「廡」係「廡」字之異體，故以「廡」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 3846 的「壻」字，Unicode 字型呈現為「壻」，實際上仍為同一字。
- 3、待造字：Unicode 及漢字構形資料庫皆未收錄的舊漢籍造字，正在等待補造字中，所以「造字」欄空白無法看到字形。如編號 5749 的「片𠂔庸」。

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 978 的「戩」字，Unicode 編碼為 4CD2，由於 Windows XP 並不支援，仍須改用構字式「戩𠂔鳥」。

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、「新校搜神記.xml」發現之未確定字形「●」共 16 處，修正情形詳列附件二，附件二「新校搜神記.xml」「●」修改表體例說明如下：

- 甲、檔案原文摘錄：摘錄檔案(新校搜神記.xml)中包含「●」未確定字形的文句段落，其中紅字底線的部分，為需要手動取代的原文。
- 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。
- 丙、備註：修改後的字形若屬於「待造字」，則於此欄標示。

附件一、「新校搜神記.xml」造字分析表

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
978	戩	3	3	8F63	4CD2		戩△鳥	
1000		1	1	8F79			王△巳	玨，異體字問題。
1039	稟	1	1	8FC2	69C0	稟		
1530	紉	1	1	92D6	7CFA	紉		
1614		1	1	936B			玨△耳	形工几耳□，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
1656	墟	1	1	93B7	380A		山 厶 虛	
1919	僑	1	1	9562	2032F		亻 厶 离	
2046	踪	1	1	9644	8E2A	踪		
2287	蝮	1	1	97BA	2736A		虫 厶 禺	
2602	鰐	3	3	99BB	9C10	鰐		
2731	窠	4	4	9A7D			宀 厶 隹	
2828	蝸	1	1	9B41	8744	蝸		
2874	餈	1	1	9B6F	9939	餈		
3095	鈞	1	1	9CD1	91F0	鈞		
3787	冢	51	51	8152	51A2	冢		
3789	冲	1	1	8154	51B2	冲		沖，異體字問題。
3791	况	1	1	8156	51B5	况		況，異體字問題。
3805	勅	1	1	8164	52C5	勅		敕，異體字問題。
3814	却	1	1	816D	5374	却		卻，異體字問題。
3819	廝	1	1	8172	53AE	廝		廝，異體字問題。
3830	咤	2	2	817D	54A4	咤		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
3834	嗶	2	2	81A3	5637	嗶		
3836	喻	1	1	81A5	564F	喻		
3840	坂	1	1	81A9	5742	坂		
3846	壻	6	6	81AF	58FB	壻		Unicode 字型差異。
3850	姪	2	2	81B3	59D9	姪		
3858	寃	7	7	81BB	5BC3	寃		寃，異體字問題。
3883	徧	1	1	81D4	5FA7	徧		
3910	擘	1	1	81EF	64E5	擘		
3931	椀	1	1	8245	6900	椀		
3974	牀	9	9	8270	7240	牀		
3978	猪	5	5	8274	732A	猪		猪，異體字問題。
3980	猿	2	2	8276	7328	猿		
3984	璫	1	1	827A	7460	璫		
3985		1	1	827B			𠄎王圭母𠄎	璫，異體字問題。
4016	着	1	1	82BC	7740	着		著，異體字問題。
4017	蟹	1	1	82BD	880F	蟹		蟹，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4032	秣	2	2	82CC	2578A		禾𠂔未	秣，異體字問題。
4038	竈	11	11	82D2	7AC8	竈		
4044	筴	1	1	82D8	7B6F	筴		
4045	箒	1	1	82D9	7B92	箒		
4080	羣	20	20	82FC	7FA3	羣		群，異體字問題。
4090	脚	2	2	8347	811A	脚		腳，異體字問題。
4093	舐	3	3	834A	445B		舌𠂔氏	舐，異體字問題。
4100	苽	1	1	8351	82FD	苽		
4128	螫	1	1	836D	87DA	螫		
4132		27	27	8371			血𠂔𠂔	眾，異體字問題。
4134	衛	13	13	8373	885E	衛		衛，異體字問題。
4139	袴	3	3	8378	88B4	袴		
4146	覩	5	5	83A1	89A9	覩		
4176	輒	4	4	83BF	8F19	輒		輒，異體字問題。
4184	迹	5	5	83C7	8FF9	迹		跡，異體字問題。
4203	隣	2	2	83DA	96A3	隣		鄰，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4237	髡	1	1	83FC	9AE0	髡		髡，異體字問題。
4242	鬪	18	18	8442	9B2D	鬪		
4256	鷄	1	1	8450	9DC4	鷄		雞，異體字問題。
4259	麇	1	1	8453	9E95	麇		Unicode 字型差異。
4260	麇	4	4	8454	9E9E	麇		
4267	鼈	4	4	845B	9F08	鼈		
4342	倮	1	1	84C8	502E	倮		
4347	蟻	1	1	84CD	27445		虫厶越	
4366	邾	1	1	84E0	48BC		共厶卩	
4460	勅	13	13	85A1	52D1	勅		
4469	勺	1	1	85AA	5304	勺		
4495	麇	1	1	85C4	9EA8	麇		
4554	唬	1	1	8640	55C1	唬		
4768	妬	3	3	8779	59AC	妬		妒，異體字問題。
4904	剋	2	2	8864	5C05	剋		剋，異體字問題。
5249	蝻	1	1	8AA5	873D	蝻		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
5292	粵	1	1	8AD0	65C9	粵		
5305	晒	9	9	8ADD	661E	晒		
5471	契	1	1	8BE6	22346		契	
5478	飴	3	3	8BED	98F0	飴		
5482	璽	1	1	8BF1	8812	璽		
5549	藁	1	1	8C75	85F3	藁		
5551	賣	1	1	8C77	8CF7	賣		
5583	殮	1	1	8CB9	98F1	殮		殮，異體字問題。
5587	歛	6	6	8CBD	6B58	歛		
5651	難	1	1	8CFD	3E10		難△庸	
5705	廐	1	1	8D74	5ED0	廐		廐，異體字問題。
5706	敝	1	1	8D75	654D	敝		敝，異體字問題。
5749		1	1	8DC2			片△庸	待造字。
5809	踰	1	1	8DFE	8E79	踰		

附件二、「新校搜神記.xml」 「●」修改表

檔案原文摘錄	手動取代結果	備註
出薯●子三枚	出薯 ^艹 △豫子三枚	待造字
立而為蚘●	立而為蚘虫△晃	待造字
過田●，數日之後	過田啜，數日之後	
稻之為●也	稻之為加△虫也	
「●國」「馬化」	「猴國」「馬化」	
四五丈，能●相喚	四五丈，能口△叢相喚	
●虫△禹	虫△敦虫△禹	
一名「●蠋，」	一名「虫△則蠋，」	
名曰「螺●，」	名曰「螺虫△羸，」	
今世謂「●●，」	今世謂「虫△因雍△虫，」	
曰：「●！●！」宜死	曰：「口△有！口△有！」宜死	
江夏●賈和姊病	江夏覃△卩賈和姊病	
不忍之，然猶●之	不忍之，然猶斫之	
●目，磋齒	瞠目，磋齒	