

《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
中央研究院語言所文獻語料小組
2009/02/12 丁致伶

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 978 的「戩」字，Unicode 編碼是 4CD2，由於 XP 並不支援，仍需使用構字式「戩△鳥」。

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：

《新校搜神記》使用舊版造字 82 個，字頻 300 次，這 82 個造字中，72 個可轉成 Windows XP 能顯示的字，字頻 283 次；另外 10 個字必須轉成構字式，字頻 17 次。

轉碼完成製作轉碼分析表，請參考附件一《新校搜神記》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 4134 的「衛」字，是「衛」的異體字，以標準字「衛」取代。
2. Unicode 字型呈現差異：Unicode 字型與舊版造字有些微差異，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 3978 的「猪」字，Unicode 字型呈現為「猪」，實際上仍為同一字。

附件一、《新校搜神記》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
978	戡	3	8F63	4CD2		戡△鳥	
1000		1	8F79			王△巳	玓，異體字問題。
1039	稟	1	8FC2	69C0	稟		
1530	紕	1	92D6	7CFA	紕		
1614	聳	1	936B	8053	聳		
1656	墟	1	93B7	380A		山△虛	
1919	儻	1	9562	2032F		亻△离	
2046	踪	1	9644	8E2A	踪		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
2287	蝮	1	97BA	2736A		虫 𧈧 禺	
2602	鰐	3	99BB	9C10	鰐		
2731	霍	4	9A7D			宀 会 隹	
2828	蝸	1	9B41	8744	蝸		
2874	餹	1	9B6F	9939	餹		
3095	鈞	1	9CD1	91F0	鈞		
3787	豕	51	8152	51A2	豕		
3789	冲	1	8154	51B2	冲		沖，異體字問題。
3791	况	1	8156	51B5	况		況，異體字問題。
3805	勅	1	8164	52C5	勅		
3819	廝	1	8172	53AE	廝		廝，異體字問題。
3830	咤	2	817D	54A4	咤		
3834	嗥	2	81A3	5637	嗥		
3836	喻	1	81A5	564F	喻		
3840	坂	1	81A9	5742	坂		
3846	壻	6	81AF	58FB	壻		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3850	姪	2	81B3	59D9	姪		
3858	寃	7	81BB	5BC3	寃		冤，異體字問題。
3883	徧	1	81D4	5FA7	徧		
3910	擘	1	81EF	64E5	擘		
3931	椀	1	8245	6900	椀		
3974	牀	9	8270	7240	牀		
3978	猪	5	8274	732A	猪		Unicode 字型呈現差異。
3980	獾	2	8276	7328	獾		
3984	瑠	1	827A	7460	瑠		
3985	璫	1	827B	7447	璫		Unicode 字型呈現差異。
4016	着	1	82BC	7740	着		
4017	螻	1	82BD	880F	螻		
4032	秣	2	82CC	2578A		禾  未	秣，異體字問題。
4038	竈	11	82D2	7AC8	竈		
4044	筋	1	82D8	7B6F	筋		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4045	箒	1	82D9	7B92	箒		
4080	羣	19	82FC	7FA3	羣		
4090	脚	2	8347	811A	脚		
4093	舐	3	834A	445B		舌 𠂔 氏	
4100	苾	1	8351	82FD	苾		
4128	螫	1	836D	87DA	螫		
4132		27	8371			血 𠂔 𠂔	眾，異體字問題。
4134	衛	12	8373	885E	衛		衛，異體字問題。
4139	袴	3	8378	88B4	袴		
4146	覩	4	83A1	89A9	覩		睹，異體字問題。
4176	輒	4	83BF	8F19	輒		
4184	迹	4	83C7	8FF9	迹		
4203	隣	2	83DA	96A3	隣		鄰，異體字問題。
4237	髡	1	83FC	9AE0	髡		
4242	鬪	18	8442	9B2D	鬪		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4256	鷄	1	8450	9DC4	鷄		雞，異體字問題。
4259	麇	1	8453	9E95	麇		
4260	麇	4	8454	9E9E	麇		
4267	鼈	4	845B	9F08	鼈		
4342	倮	1	84C8	502E	倮		
4347	蟻	1	84CD	27445		虫厶越	
4366	邾	1	84E0	48BC		共厶卩	
4460	勅	13	85A1	52D1	勅		
4469	勻	1	85AA	5304	勻		
4495	𪗇	1	85C4	9EA8	𪗇		Unicode 字型呈現差異。
4554	唬	1	8640	55C1	唬		
4768	妬	3	8779	59AC	妬		妒，異體字問題。
4904	尅	2	8864	5C05	尅		
5249	蝻	1	8AA5	873D	蝻		
5292	粵	1	8AD0	65C9	粵		
5305	晒	9	8ADD	661E	晒		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5471	契	1	8BE6	22346		契	
5478	飴	3	8BED	4D77	飴		
5482	璽	1	8BF1	8812	璽		
5549	藁	1	8C75	85F3	藁		
5551	賈	1	8C77	8CF7	賈		
5583	殮	1	8CB9	98F1	殮		
5587	歛	6	8CBD	6B58	歛		
5651	難	1	8CFD	3E10		難𠄎...	
5705	廐	1	8D74	5ED0	廐		廐，異體字問題。
5706	敘	1	8D75	654D	敘		敘，異體字問題。
5749	牖	1	8DC2	7257	牖		
5809	踰	1	8DFE	8E79	踰		

二、手動取代表說明：

《新校搜神記》檔案中出現「●」字形，共計16字，比對原書後以Unicode字形或構字式手動取代，並製作「●」字手動取代表，請參考附件二。欄位說明如下：

1. 頁碼：「●」字形所在位置之原書頁碼。

2. 檔案原文摘錄：摘錄《新校搜神記》檔案有「●」字形的文句。
3. 原書字形：「●」在原書中之字形，以此字取代檔案原文的「●」。
4. 備註：記錄修改相同詞句的次數或其他事項。

附件二、《新校搜神記》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 10	出薯●子三枚	𦉳 𧈧 豫	待造字
p. 41	蚺●，此天之常	虫 𧈧 晃	待造字
p. 58	過田●，數日	𧈧	
p. 89	稻之為●也	加 𧈧 虫	
p. 93	率皆是「●國」	猓	
p. 94	能●相喚	口 𧈧 叢	
p. 99	南方有蟲，名「●	虫 𧈧 敦	
p. 99	一名「●蠋，」	虫 𧈧 則	
p. 99	名曰「螺●，」	虫 𧈧 羸	
p. 99	今世謂「●●，」	虫 𧈧 因 雍 𧈧 虫	
p. 106	曰：「●！●！」宜死	𧈧 𧈧	
p. 111	江夏●賈和姊病	覃 𧈧 卩	
p. 128	然猶●之	石 𧈧 斤	改為斫
p. 141	與大賢語，●目	𧈧	