

# 《赤水玄珠全集》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  
2008/3/19 陳建安 製作

一、《赤水玄珠全集》一書(赤水玄珠全集.xml)使用舊版造字 116 個，字頻 275 次，詳如附件一。這 116 個造字中，94 個可轉成 Windows XP 能顯示的字，字頻 238 次；另外 22 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 674 的「乚」字，原字為「匕」。另外，編號 3353 及 3618，原書並無字形，手動加以刪除。

2、校對問題，新舊版缺字並存：漢籍在舊轉新的過程中，以人工修改方式取代舊漢籍造字，雖然修改的結果是正確的，但是只修改到部分造字，沒有全部修改到，導致舊漢籍造字在新版漢籍檔案中，一部分已被取代成 Unicode 或構字式，另一部分仍為舊漢籍造字的並存情形。處理方式為配合漢籍工作室的修改，將沒被修改到的舊漢籍造字，以對應的 Unicode 或構字式取代。如編號 1147，舊漢籍造字為「𠃉穴矣欠◻」，比對 txt 檔及 xml 檔後，發現原先在 txt 檔使用舊漢籍造字「𠃉穴矣欠◻」者，在 xml 檔中部分已人工修改為「竅」字，故配合修改，將編號 1147「𠃉穴矣欠◻」字，手動取代為

「窳」字。

- 3、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 14 的「𠄎彳幺月乚◻」係「胤」字之異體，故以「胤」字取代。
- 4、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 3832 的「嘅」字，Unicode 字型呈現為「嘅」，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 1334 的「眇」字，Unicode 編碼為 26548，由於 Windows XP 並不支援，仍須使用構字式「耳厶少」。

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、txt 檔與 xml 檔頻次不同且造字分析表「備註」欄有備註問題情況者，詳細取代內容可參考附件二的手動取代表。附件二的手動取代表的欄位說明如下：

- 甲、編號：Big5 造字空間為 6217 個，編號由 1 到 6217。
- 乙、取代：當實際使用字形與造字不同時，在此標出。
- 丙、檔案原文摘錄：摘錄檔案(赤水玄珠全集.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。文句後方標示(全部取代)者，表示執行全部取代的動作，不一一列舉所有原文。
- 丁、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。文句後方標示(全部取代)者，表示執行全部取代的動作，不一一列舉所有原文。

附件一、《赤水玄珠全集》造字分析表

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
14	胤	1	1	FA4D	38E7		𠃉彳幺月乚	胤，異體字問題。
602	氷	1	1	FDE4			氷	
639	彳	1	1	FE4A			彳	
674	匕	3	3	FE6D			匕	匕，校對問題，舊版漢籍錯字，可用程式全部取代。
713	尸	1	1	FEB6			尸	
752	𠃉	1	1	FEDD			𠃉	
912	鷓	1	1	8EE0	9DC0	鷓		
938	燮	3	3	8EFA	7215	燮		
1147		2	1	906F			𠃉穴矣欠	窾，校對問題，新舊版缺字並存。
1151	劓	1	1	9073	5273	劓		
1239		1	1	90ED			𠃉豆	豎，異體字問題。
1243	桐	1	1	90F1	79F1	桐		
1321	確	3	3	91A2	7936	確		
1334	眇	1	1	91AF	26548		耳少	
1416	埃	3	3	9242	7AE2	埃		
1491	擷	2	2	92AF	6527	擷		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
1544	絙	1	1	92E4	7D59	絙		
1642	脉	1	1	93A9	8109	脉		
1658	函	1	1	93B9	5705	函		
1660	賺	4	4	93BB	8181	賺		
1664	腸	1	1	93BF	8193	腸		
1905	棍	4	4	9554	88E9	棍		
1951	味	1	1	95A4	20C0C		口厶末	味，校對問題，舊版漢籍錯字，可用程式全部取代。
1999	贊	1	1	95D4	8CDB	贊		贊，異體字問題。
2042	疏	1	1	9640	47FD		足厶充	疏，異體字問題。
2046	踪	2	2	9644	8E2A	踪		
2140	迫	3	3	96C4	5EF9	迫		迫，異體字問題。
2142	冢	1	1	96C6	7758	冢		Unicode 呈現差異。
2338	苜	1	1	97ED	82E9	苜		
2374	冢	4	4	9852	51A1	冢		冢，異體字問題。
2524	饞	1	1	994B	994D	饞		
2649	賺	2	2	99EA	28891		酉厶兼	
2885	叫	2	2	9B7A	544C	叫		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
3064	鯛	1	1	9CB2	9C02	鯛		
3161	壘	2	2	9D54	58DC	壘		
3269	憑	1	1	9DE2	51F4	憑		
3353		1	1	9E77				校對問題，舊版漢籍錯字，無字形，手動刪除。
3618		1	1	A046				校對問題，舊版漢籍錯字，無字形，手動刪除。
3787	冢	2	2	8152	51A2	冢		
3796	決	3	3	815B	51B3	決		決，異體字問題。
3801	剗	1	1	8160	5257	剗		
3805	勅	3	3	8164	52C5	勅		
3820	廈	1	1	8173	53A6	廈		廈，異體字問題。
3824	叙	6	4	8177	53D9	叙		敘，校對問題，新舊版缺字並存。
3825	疊	4	4	8178	53E0	疊		疊，異體字問題。
3832	嘅	3	3	81A1	5605	嘅		Unicode 呈現差異。
3843	堦	2	2	81AC	5826	堦		
3909	擡	3	3	81EE	64E1	擡		
3927	柏	9	6	8241	6855	柏		柏，校對問題，新舊版缺字並存。
3931	椀	1	1	8245	6900	椀		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
3935	槩	1	1	8249	69E9	槩		概，異體字問題。
3951	湏	1	1	8259	6E4F	湏		
3974	牀	14	14	8270	7240	牀		
3977	狗	1	1	8273	72E5	狗		
3996	疎	17	16	82A8	758E	疎		疎，校對問題，新舊版缺字並存。
4002	瘡	1	1	82AE	7607	瘡		
4006	癢	2	2	82B2	764F	癢		
4008	癯	1	1	82B4	766F	癯		
4010	臯	1	1	82B6	7690	臯		Unicode 呈現差異。
4015	眦	3	3	82BB	7726	眦		
4033	稽	1	1	82CD	7A2D	稽		
4038	竈	1	1	82D2	7AC8	竈		
4041	豎	6	6	82D5	7AEA	豎		
4042	笋	3	3	82D6	7B0B	笋		
4044	筋	12	12	82D8	7B6F	筋		
4068	綫	1	1	82F0	7DAB	綫		線，異體字問題。
4069	綑	1	1	82F1	7DB3	綑		
4091	臯	3	3	8348	81EF	臯		Unicode 呈現差異。

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4109	葛	1	1	835A	84AD	葛		
4139	袴	2	2	8378	88B4	袴		
4140	袵	2	2	8379	88B5	袵		
4144	霸	2	2	837D	8987	霸		Unicode 呈現差異。
4146	覩	5	5	83A1	89A9	覩		
4158	譌	1	1	83AD	8B4C	譌		Unicode 呈現差異。
4169	趁	1	1	83B8	8DA6	趁		
4173	躄	5	5	83BC	8EAD	躄		
4178	輓	1	1	83C1	8F2D	輓		
4186	遡	1	1	83C9	9061	遡		
4195	鉢	1	1	83D2	9262	鉢		
4203	隣	4	4	83DA	96A3	隣		鄰，異體字問題。
4223	顛	1	1	83EE	984B	顛		
4226	飡	2	2	83F1	98E1	飡		
4242	鬪	2	2	8442	9B2D	鬪		Unicode 呈現差異。
4247	鮎	2	2	8447	9B8E	鮎		
4251	鰕	5	5	844B	9C15	鰕		
4262	麪	3	3	8456	9EAA	麪		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4267	鼈	2	2	845B	9F08	鼈		
4278	亘	2	2	8466	4E98	亘		
4299	炳	3	3	847B	242F2		火𠂇𠂇	炳，異體字問題。
4307	罵	6	6	84A5	99E1	罵		罵，異體字問題。
4308	劫	2	2	84A6	5227	劫		劫，異體字問題。
4400	兜	4	4	8543	5160	兜		
4428	蹶	1	1	855F	8E70	蹶		
4442	潛	1	1	856D	6F98	潛		Unicode 呈現差異。
4546	鷓	1	1	85F7	9D76	鷓		
4572	嘍	4	4	8652	20DFA		口𠂇勃	
4778	鞞	2	2	87A5	97B4	鞞		
4803	腭	8	6	87BE	816D	腭		腭，校對問題，新舊版缺字並存。
4805	瀆	3	3	87C0	51DF	瀆		瀆，異體字問題。
4822	尪	1	1	87D1	5C29	尪		
4910		6	6	886A			彡𠂇𠂇	淫，異體字問題。
5013	彡	1	1	88F3			彡	
5119	博	1	1	89C0	613D	博		博，異體字問題。
5129	杵	1	1	89CA	676B	杵		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
5173	漱	1	1	89F6	6F44	漱		
5193	號	2	2	8A4B	8662	號		
5200		1	1	8A52			𠄎ノ門米大◻	奧，異體字問題。
5300	鬪	3	3	8AD8	9B2A	鬪		
5312		1	1	8AE4			𠄎女ノ岡比◻	媿，異體字問題。
5408	絳	3	2	8BA7	7D4D	絳		絳，校對問題，新舊版缺字並存。
5448	鬪	6	6	8BCF	9AC3	鬪		
5555	迥	1	1	8C7B	9008	迥		
5590	朶	1	1	8CC0	6736	朶		朶，異體字問題。
5591	杞	2	2	8CC1	233CC		木△巳	杞，異體字問題。
5595	朶	1	1	8CC5	579C	朶		
5783	教	2	2	8DE4	3A8D		孛△攵	

附件二、《赤水玄珠全集》手動取代表

編號	取代	檔案原文摘錄	手動取代結果
1147	窳	痘疹胎產皆中__	痘疹胎產皆中窳
3353		胃脘當心而痛， <u>乱</u>	胃脘當心而痛，
3618		且痛， <u>俞</u>	且痛，
3824	敘	<u>甄</u> （全部取代）	<u>敘</u> （全部取代）
3927		<u>將</u> （全部取代）	<u>柏</u> （全部取代）
3996		<u>许</u> （全部取代）	<u>疎</u> （全部取代）
4803		<u>黍</u> （全部取代）	<u>腭</u> （全部取代）
5408		<u>亲</u> （全部取代）	<u>絳</u> （全部取代）