

《鐵圍山叢談》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
2008/6/24 陳維君 製作

一、《鐵圍山叢談》一書(鐵圍山叢談.xml)使用舊版造字 77 個，字頻 789 次，詳如附件一。這 77 個造字中，64 個可轉成 Windows XP 能顯示的字，字頻 726 次；另外 13 個字必須轉成構字式，字頻 63 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 4132 的「血𩇛𩇛」係「眾」字之異體，故以「眾」字取代。又編號 3966 的「烟」字，係「煙」字之簡體字，亦以標準字的「煙」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 4158 的「譌」字，Unicode 字型呈現為「譌」，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 4415 的「藿」字，Unicode 編碼為 27191，由於 Windows XP 並不支援，仍須使用構字式「𩇛𩇛𩇛」。

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、《鐵圍山叢談》造字分析表

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
296	𡗗	2	2	FBEC			𡗗	
297	𡗘	2	2	FBED			𡗘	
299	𡗙	1	1	FBEF			𡗙	
300	𡗚	1	1	FBF0			𡗚	
348	与	1	1	FC61			与	
842	𡗛	2	2	8E78	3AB7		𡗛	
860	𡗜	1	1	8EAC	890F	𡗜		
893	𡗝	1	1	8ECD	5CE9	𡗝		峨，異體字問題。
1048	𡗞	1	1	8FCB	981F	𡗞		
1067	𡗟	7	7	8FDE	68CA	𡗟		棋，異體字問題。
1416	埃	1	1	9242	7AE2	埃		
1544	𡗠	4	4	92E4	7D59	𡗠		
1893	袞	3	3	9548	886E	袞		袞，異體字問題。
1981	𡗡	1	1	95C2	8B90	𡗡		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
2046	踪	1	1	9644	8E2A	踪		
2213	鈇	1	1	974E	9277	鈇		
3152	錢	1	1	9D4B	921B	錢		
3801	剗	1	1	8160	5257	剗		
3802	劓	1	1	8161	529A	劓		
3809	匱	5	5	8168	5333	匱		
3846	墻	1	1	81AF	58FB	墻		Unicode 呈現差異。
3852	媼	1	1	81B5	5A63	媼		
3860	尠	3	3	81BD	5C20	尠		
3883	徧	7	7	81D4	5FA7	徧		
3931	椀	1	1	8245	6900	椀		
3935	槩	5	5	8249	69E9	槩		概，異體字問題。
3965	栽	1	1	8267	70D6	栽		
3966	烟	4	4	8268	70DF	烟		煙，異體字問題。
3968	煨	1	1	826A	7145	煨		
3974	牀	4	4	8270	7240	牀		
3989	璿	3	3	82A1	74C8	璿		
3996	疎	4	4	82A8	758E	疎		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
4038	竈	1	1	82D2	7AC8	竈		
4040	竝	253	253	82D4	7ADD	竝		
4044	筋	1	1	82D8	7B6F	筋		
4063	藥	1	1	82EB	7CF5	藥		
4066	綺	1	1	82EE	7D5D	綺		
4080	羣	41	41	82FC	7FA3	羣		群，異體字問題。
4118	藥	9	9	8363	8602	藥		
4131	衃	1	1	8370	8842	衃		
4132		44	44	8371			血衃衃	眾，異體字問題。
4134	衛	1	1	8373	885E	衛		衛，異體字問題。
4138	表	2	2	8377	88A0	表		
4139	袴	2	2	8378	88B4	袴		
4142	裛	1	1	837B	8943	裛		
4146	覩	3	3	83A1	89A9	覩		
4158	譌	6	6	83AD	8B4C	譌		Unicode 呈現差異。
4165	獾	2	2	83B4	8C9B	獾		
4184	迹	2	2	83C7	8FF9	迹		
4219	頰	1	1	83EA	982E	頰		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
4222	頤	1	1	83ED	9847	頤		
4237	髡	9	9	83FC	9AE0	髡		髡，異體字問題。
4242	鬪	3	3	8442	9B2D	鬪		Unicode 呈現差異。
4256	鷄	1	1	8450	9DC4	鷄		
4267	鼈	2	2	845B	9F08	鼈		
4275	义	1	1	8463	4E49	义		
4342	倮	2	2	84C8	502E	倮		
4415	藿	2	2	8552	27191		艹 艹 藿	
4423	醫	1	1	855A	81CB	醫		
4461	勛	1	1	85A2	52D4	勛		
4469	勺	2	2	85AA	5304	勺		
4482	廔	7	7	85B7	53AA	廔		廔，異體字問題。
4507	吴	289	289	85D0	5434	吴		吳，異體字問題。
4608	嚮	1	1	8676	7581	嚮		
4630		1	1	86AE			勹 匕 示 欠 勹	款，異體字問題。
4902	兗	1	1	8862	5156	兗		兗，異體字問題。
5008	幞	1	1	88EE	5E5E	幞		
5013	汜	1	1	88F3			汜	

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
5075	彳	1	1	8972			彳	
5302	昏	4	4	8ADA	662C	昏		昏，異體字問題。
5344	磬	2	2	8B45	55F8	磬		嗽，異體字問題。
5537	离	2	2	8C69	79BC	离		
5592	桌	1	1	8CC2	3B9A		桌	
5685	懣	4	4	8D60	22960		忄 ㄩ 雙	
5706	敘	1	1	8D75	654D	敘		敘，異體字問題。
5797	灑	1	1	8DF2	7054	灑		
5854	幺	2	2	C6CD	5E7A	幺		