

《鄭思肖集》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

2008/4/15 陳建安 製作

一、《鄭思肖集》一書(鄭思肖集.xml)使用舊版造字 160 個，字頻 891 次，詳如附件一。這 160 個造字中，137 個可轉成 Windows XP 能顯示的字，字頻 755 次；另外 23 個字必須轉成構字式，字頻 136 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 3094 的「衤𠂔異」字，原字為「禪」。
- 2、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「血𩺰𩺰」係「眾」字之異體，故以「眾」字取代。又編號 4068 的「綫」字，係「線」字之簡體字，亦以標準字的「線」字取代。
- 3、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 2234 的「𪗇」字，Unicode 編碼為 38B8，由於 Windows XP 並不支援，仍須使用構字式「𪗇 𪗇」。

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、《鄭思肖集》造字分析表

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
296	𪗇	3	3	FBEC			𪗇	
297	𪗇	7	7	FBED			𪗇	
298	𪗇	1	1	FBEE			𪗇	
685	𪗇	1	1	FE78			𪗇	
770	𪗇	1	1	FEEF			𪗇	
895	𪗇	1	1	8ECF	71A2	𪗇		
896	𪗇	1	1	8ED0	6AE9	𪗇		
969	𪗇	1	1	8F5A	7250	𪗇		
1010	𪗇	5	5	8FA5	73CF	𪗇		
1024	𪗇	1	1	8FB3	8675	𪗇		
1056	𪗇	1	1	8FD3	9641	𪗇		
1151	𪗇	2	2	9073	5273	𪗇		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
1205	眎	1	1	90CB	770E	眎		
1354	褸	2	2	91C3	79A9	褸		
1400	窓	1	1	91F1	7A93	窓		
1437	舫	1	1	9257	7265	舫		
1491	擲	1	1	92AF	6527	擲		
1530	紂	3	3	92D6	7CFA	紂		
1565	繇	2	2	92F9	7DDC	繇		
1642	脉	7	7	93A9	8109	脉		
1646	脬	2	2	93AD	8117	脬		
1660	賺	1	1	93BB	8181	賺		
1700		1	1	93E3			𣎵 𣎵 𣎵 𣎵 心	𣎵 𣎵 𣎵 心 𣎵，異體字問題。
1875	割	1	1	94F5	5268	割		
1885	蠡	1	1	9540	882D	蠡		
1893	袞	1	1	9548	886E	袞		
1912	襍	1	1	955B	894D	襍		
1981	讐	7	7	95C2	8B90	讐		
2044	踈	2	2	9642	8E08	踈		
2157	酎	1	1	96D5	9167	酎		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
2234	𠂔	2	2	9763	38B8		𠂔 𠂔 𠂔	
2409	𨵿	1	1	9875	95A7	𨵿		𨵿，異體字問題。
2448	隼	1	1	98BE	96BD	隼		隼，異體字問題。
2602	鰐	3	3	99BB	9C10	鰐		
2651	鸛	1	1	99EC	9E1C	鸛		
2742	抄	1	1	9AAA	6331	抄		
3028	𠂔	1	1	9C6C			𠂔	
3094		9	9	9CD0			𠂔 𠂔 異	𠂔，校對問題，舊版漢籍錯字，可用程式全部取代。
3244	𦰩	1	1	9DC9	686C	𦰩		
3269	𦰪	1	1	9DE2	51F4	𦰪		
3796	決	3	3	815B	51B3	決		決，異體字問題。
3799	𠂔	7	7	815E	5226	𠂔		
3801	𠂔	1	1	8160	5257	𠂔		
3804	効	2	2	8163	52B9	効		
3805	勅	8	8	8164	52C5	勅		
3807	驗	8	8	8166	9A10	驗		
3814	却	51	51	816D	5374	却		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
3816	厓	3	3	816F	5393	厓		
3824	叙	16	16	8177	53D9	叙		敘，異體字問題。
3829	咏	7	7	817C	548F	咏		
3830	咤	1	1	817D	54A4	咤		
3846	壻	3	3	81AF	58FB	壻		Unicode 呈現差異。
3858	寃	20	20	81BB	5BC3	寃		冤，異體字問題。
3864	峯	14	14	81C1	5CEF	峯		峰，異體字問題。
3868	卮	1	1	81C5	5DF5	卮		卮，異體字問題。
3883	徧	10	10	81D4	5FA7	徧		
3884	忽	2	2	81D5	6031	忽		
3891	慙	3	3	81DC	6159	慙		慚，異體字問題。
3930	憵	2	2	8244	6901	憵		
3931	椀	1	1	8245	6900	椀		
3932	椶	1	1	8246	6936	椶		
3935	槩	1	1	8249	69E9	槩		概，異體字問題。
3944	氷	3	3	8252	6C37	氷		冰，異體字問題。
3945		4	4	8253			汊 𠂇义	汊，異體字問題。
3963	炁	4	4	8265	7081	炁		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
3968	煨	2	2	826A	7145	煨		
3974	牀	4	4	8270	7240	牀		
3978	猪	1	1	8274	732A	猪		Unicode 呈現差異。
3989	璦	1	1	82A1	74C8	璦		
3990	甄	2	2	82A2	750E	甄		
3996	疎	9	9	82A8	758E	疎		
4008	癰	3	3	82B4	766F	癰		
4016	着	15	15	82BC	7740	着		著，異體字問題。
4025	礪	2	2	82C5	7934	礪		
4028	稟	2	2	82C8	7980	稟		稟，異體字問題。
4034	穉	1	1	82CE	7A49	穉		
4035	穽	2	2	82CF	7A7D	穽		
4038	竈	2	2	82D2	7AC8	竈		
4052	簾	1	1	82E0	7C34	簾		
4056	葬	33	33	82E4	7CA6	葬		
4068	綫	8	8	82F0	7DAB	綫		線，異體字問題。
4080	羣	40	40	82FC	7FA3	羣		群，異體字問題。
4090	脚	9	9	8347	811A	脚		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4091	皐	13	13	8348	81EF	皐		Unicode 呈現差異。
4097	芑	1	1	834E			艹厶巳	芑，異體字問題。
4107	葱	3	3	8358	8471	葱		
4115	藁	3	3	8360	85C1	藁		
4124	𧈧	1	1	8369	8771	𧈧		
4132		88	88	8371			血厶𠂇	眾，異體字問題。
4134	衛	4	4	8373	885E	衛		衛，異體字問題。
4140	衽	3	3	8379	88B5	衽		
4142	褰	1	1	837B	8943	褰		
4146	覩	16	16	83A1	89A9	覩		
4158	譌	1	1	83AD	8B4C	譌		Unicode 呈現差異。
4167	賁	1	1	83B6	8CEB	賁		
4173	𧈧	1	1	83BC	8EAD	𧈧		
4176	輒	4	4	83BF	8F19	輒		
4184	迹	30	30	83C7	8FF9	迹		
4186	遡	2	2	83C9	9061	遡		
4193	鈎	4	4	83D0	920E	鈎		
4195	鉢	1	1	83D2	9262	鉢		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4203	隣	1	1	83DA	96A3	隣		鄰，異體字問題。
4213	韡	1	1	83E4	97BE	韡		
4220	穎	3	3	83EB	9834	穎		穎，異體字問題。
4226	飡	1	1	83F1	98E1	飡		
4232	鷺	8	8	83F7	96B2	鷺		鷺，異體字問題。
4236	瞢	1	1	83FB	77D2	瞢		
4237	髡	2	2	83FC	9AE0	髡		髡，異體字問題。
4240	髯	4	4	8440	9AF4	髯		
4242	鬪	3	3	8442	9B2D	鬪		Unicode 呈現差異。
4256	鷄	1	1	8450	9DC4	鷄		
4278	亘	8	8	8466	4E98	亘		
4307	罵	57	57	84A5	99E1	罵		罵，異體字問題。
4308	劫	18	18	84A6	5227	劫		劫，異體字問題。
4309	刼	22	22	84A7	523C	刼		刼，異體字問題。
4348	惚	1	1	84CE	506C	惚		
4437	廩	2	2	8568	5EEA	廩		廩，異體字問題。
4450	𪔐	1	1	8575	524F	𪔐		𪔐，異體字問題。
4460	勅	1	1	85A1	52D1	勅		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4495	尅	1	1	85C4	9EA8	尅		
4515	呪	12	12	85D8	546A	呪		
4533	唸	2	2	85EA	552B	唸		
4603	憇	1	1	8671	6187	憇		
4608	𪔐	1	1	8676	7581	𪔐		
4702	塲	1	1	86F6	5872	塲		
4718	𪔑	1	1	8747	340C		𪔑	
4738	睺	1	1	875B	777A	睺		
4750	𪔒	1	1	8767			𪔒	升，校對問題，舊版漢籍錯字，可用程式全部取代。
4768	妬	7	7	8779	59AC	妬		妒，異體字問題。
4785	𪔓	3	3	87AC	8D0B	𪔓		𪔓，異體字問題。
4804	𪔔	1	1	87BF	5E42	𪔔		
4817	寶	37	37	87CC	5BF3	寶		寶，異體字問題。
4903	𪔕	1	1	8863	5C00	𪔕		
4904	𪔖	1	1	8864	5C05	𪔖		
4907	𪔗	5	5	8867	69D6	𪔗		
4932	𪔘	2	2	88A2	21DEB		山𪔘成	

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4955	龜	2	2	88B9	9F9C		龜	龜，異體字問題。
5013	汜	1	1	88F3			汜	
5019	畧	1	1	88F9			畧	略，異體字問題。
5163	憂	2	2	89EC	621E	憂		
5169	危	1	1	89F2	6239	危		
5173	漱	1	1	89F6	6F44	漱		漱，異體字問題。
5176		1	1	89F9			汜 厶 冗	沉，異體字問題。
5289	慤	1	1	8ACD	6164	慤		
5290	携	1	1	8ACE	643A	携		
5303	昱	6	6	8ADB	6630	昱		
5315	沿	2	2	8AE7	3CC2		汜 厶 公	沿，異體字問題。
5402	朮	5	5	8BA1	672F	朮		
5442	枿	1	1	8BC9	67BF	枿		
5541	蠹	1	1	8C6D	8827	蠹		
5555	迥	2	2	8C7B	9008	迥		
5565	斛	3	3	8CA7	469F		角 厶 寻	
5587	歛	1	1	8CBD	6B58	歛		
5590	朶	1	1	8CC0	6736	朶		朶，異體字問題。

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
5591	杞	2	2	8CC1	233CC		木𣎵已	杞，異體字問題。
5619	𪔐	1	1	8CDD	470C		𪔐	
5666	汚	9	9	8D4D	6C5A	汚		汚，異體字問題。
5697	𪔑	1	1	8D6C	7151	𪔑		煮，異體字問題。
5706	𪔒	57	57	8D75	654D	𪔒		敘，異體字問題。
5742	𪔓	3	3	8DBB	6F81	𪔓		