

《邵氏聞見錄》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
2008/6/20 陳維君 製作

一、《邵氏聞見錄》一書(邵氏聞見錄.xml)使用舊版造字 92 個，字頻 351 次，詳如附件一。這 92 個造字中，78 個可轉成 Windows XP 能顯示的字，字頻 282 次；另外 14 個字必須轉成構字式，字頻 69 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 3800 的「杀𠄎」係「剝」字之異體，故以「剝」字取代。又編號 3942 的「毡」字，係「氈」字之簡體字，亦以標準字的「氈」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 4242 的「鬪」字，Unicode 字型呈現為「鬪」，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 2188 的「羅」字，Unicode 編碼為 4366，由於 Windows XP 並不支援，仍須使用構字式「𠄎離」。

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、《邵氏聞見錄》造字分析表

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
297	𡗗	9	9	FBED			𡗗	
842	𡗗	2	2	8E78	3AB7		𡗗	
1026	珣	2	2	8FB5	73E6	珣		
1040	聃	1	1	8FC3	803C	聃		
1067	碁	1	1	8FDE	68CA	碁		碁，異體字問題。
1097	嶠	1	1	8FFC	37ED		山 𡗗 昏	
1133	瓌	3	3	9061	74CC	瓌		
1146	甕	2	2	906E	7506	甕		
1187	盖	1	1	90B9	76D6	盖		盖，異體字問題。
1226	罰	2	2	90E0	7F78	罰		
1945	詢	2	2	957C	8A7E	詢		
2151	𠃉	8	8	96CF	961D	𠃉		
2188	離	1	1	96F4	4366		𠃉 𡗗 離	
2524	饒	1	1	994B	994D	饒		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
2651	鸚	1	1	99EC	9E1C	鸚		
3096	歲	1	1	9CD2	5D57	歲		歲，異體字問題。
3789	冲	6	6	8154	51B2	冲		沖，異體字問題。
3796	決	1	1	815B	51B3	決		決，異體字問題。
3800		1	1	815F			杀 Δ 丿	剝，異體字問題。
3801	剝	2	2	8160	5257	剝		
3804	効	1	1	8163	52B9	効		
3805	勅	1	1	8164	52C5	勅		
3814	却	22	22	816D	5374	却		卻，異體字問題。
3819	廝	13	13	8172	53AE	廝		廝，異體字問題。
3820	廈	1	1	8173	53A6	廈		廈，異體字問題。
3831	詰	1	1	817E	5586	詰		
3840	坂	1	1	81A9	5742	坂		
3846	壻	7	7	81AF	58FB	壻		Unicode 呈現差異。
3858	寃	4	4	81BB	5BC3	寃		冤，異體字問題。
3883	徧	1	1	81D4	5FA7	徧		
3891	慙	1	1	81DC	6159	慙		慚，異體字問題。
3924	朞	1	1	81FD	671E	朞		期，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
3931	椀	3	3	8245	6900	椀		
3935	槩	4	4	8249	69E9	槩		
3936	椁	1	1	824A	6A70	椁		
3942	毡	1	1	8250	6BE1	毡		氈，異體字問題。
3974	牀	8	8	8270	7240	牀		
3996	疎	11	11	82A8	758E	疎		
4015	眦	1	1	82BB	7726	眦		
4016	着	3	3	82BC	7740	着		著，異體字問題。
4028	稟	2	2	82C8	7980	稟		稟，異體字問題。
4036	窰	1	1	82D0	7AB0	窰		
4039	竝	2	2	82D3	7ADA	竝		
4041	豎	1	1	82D5	7AEA	豎		
4043	筭	1	1	82D7	7B53	筭		筭，異體字問題。
4044	筋	1	1	82D8	7B6F	筋		
4077	罇	1	1	82F9	7F47	罇		
4078		1	1	82FA			巳𠄎升	异，異體字問題。
4080	羣	25	25	82FC	7FA3	羣		群，異體字問題。
4090	脚	1	1	8347	811A	脚		腳，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
4091	臯	1	1	8348	81EF	臯		
4132		36	36	8371			血𠄎𠄎	眾，異體字問題。
4134	衛	17	17	8373	885E	衛		衛，異體字問題。
4140	衽	1	1	8379	88B5	衽		
4145	羈	1	1	837E	898A	羈		Unicode 呈現差異。
4146	覩	7	7	83A1	89A9	覩		
4158	譌	1	1	83AD	8B4C	譌		Unicode 呈現差異。
4176	輒	7	7	83BF	8F19	輒		
4184	迹	7	7	83C7	8FF9	迹		
4193	鈎	1	1	83D0	920E	鈎		鈎，異體字問題。
4195	鉢	2	2	83D2	9262	鉢		
4220	穎	2	2	83EB	9834	穎		穎，異體字問題。
4226	飡	2	2	83F1	98E1	飡		
4242	鬪	2	2	8442	9B2D	鬪		Unicode 呈現差異。
4307	罵	2	2	84A5	99E1	罵		罵，異體字問題。
4309	劫	2	2	84A7	523C	劫		劫，異體字問題。
4442	潛	1	1	856D	6F98	潛		Unicode 呈現差異。
4448		1	1	8573			𠄎絲 厂墨 𠄎	纏，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
4460	勅	6	6	85A1	52D1	勅		
4490	叁	1	1	85BF	4E9D	叁		
4502	俎	1	1	85CB	723C	俎		
4507	吳	30	30	85D0	5434	吳		吳，異體字問題。
4603	憇	1	1	8671	6187	憇		
4813		3	3	87C8			灑灑 灑灑	灑，異體字問題。
4957	崑	2	2	88BB	5D53	崑		
5008	幞	1	1	88EE	5E5E	幞		
5119	博	2	2	89C0	613D	博		
5126	暮	6	6	89C7	8B29	暮		
5176		7	7	89F9			沉 沉	沉，異體字問題。
5218		4	4	8A64			寬 寬、寬	寬，異體字問題。
5225	邨	1	1	8A6B	460F		邨 邨	邨，異體字問題。
5290	携	3	3	8ACE	643A	携		
5325	踈	1	1	8AF1	28082		踈 踈	
5437	并	1	1	8BC4	5E77	并		并，異體字問題。
5542	鑠	1	1	8C6E	93C1	鑠		
5555	迥	1	1	8C7B	9008	迥		迥，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
5558	龕	1	1	8C7E	9E84	龕		
5580	离	4	4	8CB6	79BC	离		
5588	恡	1	1	8CBE	6061	恡		
5591	杞	1	1	8CC1	233CC		木厶巳	杞，異體字問題。
5704	穉	1	1	8D73	6C0A	穉		
5706	敘	8	8	8D75	654D	敘		敘，異體字問題。