

《舊典備徵》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
2008/7/24 葉冠孜 製作

一、《舊典備徵》一書(舊典備徵.xml)使用舊版造字 90 個，字頻 432 次，詳如附件一。這 90 個造字中，75 個可轉成 Windows XP 能顯示的字，字頻 393 次；另外 15 個字必須轉成構字式，字頻 39 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「血𠂇𠂇」係「眾」字之異體，故以「眾」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 4740 的「𦵑」字，Unicode 字型呈現為「𦵑」，實際上仍為同一字。
- 3、待造字：Unicode 及漢字構形資料庫皆未收錄的舊漢籍造字，正在等待補造字中，所以「造字」欄空白無法看到字形。如編號 817 的「火𠂇呈」。

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 1940 的「襪」字，Unicode 編碼為 4109，由於 Windows XP 並不支援，仍須使用構字式「衤𠂇我」。

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、《舊典備徵》造字分析表

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
296	𠄎	4	4	FBEC			𠄎	
297	𠄍	3	3	FBED			𠄍	
298	𠄌	2	2	FBEE			𠄌	
299	𠄋	2	2	FBEF			𠄋	
300	𠄊	2	2	FBF0			𠄊	
466	𠄉	2	2	FCF9			𠄉	
780	𠄈	3	3	FEF9			𠄈	
817		1	1	8E5F			火𠄍呈	待造字。
820	烟	1	1	8E62	70F1	烟		
893	峩	1	1	8ECD	5CE9	峩		
1010	珏	1	1	8FA5	73CF	珏		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
1025	城	2	2	8FB4	73F9	城		
1092	瑞	1	1	8FF7	747A	瑞		
1118	倅	3	3	9052	2032D		亻 厶 率	
1148		1	1	9070			𠂇 厶 汙	汙 厶 符，異體字問題。
1189	璧	2	2	90BB	24A71		殿 厶 玉	
1205	眇	1	1	90CB	770E	眇		
1226	罰	1	1	90E0	7F78	罰		
1354	禊	1	1	91C3	79A9	禊		
1360	禿	3	3	91C9	79B5	禿		
1565	繇	16	16	92F9	7DDC	繇		
1627	禍	1	1	9378	7991	禍		
1634	忬	1	1	93A1	5FEC	忬		
1876	愬	1	1	94F6	6133	愬		
1940	袂	1	1	9577	4109		衤 厶 我	
1943	詒	1	1	957A	8A65	詒		
1956	諒	2	2	95A9	8AB4	諒		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
1964	還	1	1	95B1	8B5E	還		
1966	翻	2	2	95B3	4397		喬厶羽	
2248	鑫	1	1	9771	9331	鑫		
2274	鎡	1	1	97AD	9348	鎡		
2359	鑄	2	2	9843	9408	鑄		
2434	随	1	1	98B0	968F	随		
2436	愴	2	2	98B2	61BB	愴		
2592	贓	1	1	99B1	8D1C	贓		
3780	雋	3	3	814B	5101	雋		
3789	冲	1	1	8154	51B2	冲		
3814	却	1	1	816D	5374	却		
3824	叙	1	1	8177	53D9	叙		
3842	堃	5	5	81AB	5803	堃		
3843	堦	1	1	81AC	5826	堦		
3858	寃	1	1	81BB	5BC3	寃		
3866	崑	3	3	81C3	5D10	崑		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
3888	憇	3	3	81D9	60B3	憇		
3897	憇	1	1	81E2	615C	憇		
3910	擘	1	1	81EF	64E5	擘		
3916	敷	1	1	81F5	656D	敷		
3933	架	1	1	8247	6998	架		
3969	煊	6	6	826B	714A	煊		
3976	犁	2	2	8272	7282	犁		
3987	穀	1	1	827D	7474	穀		
3988	璦	1	1	827E	7482	璦		
4068	綫	1	1	82F0	7DAB	綫		
4080	羣	11	11	82FC	7FA3	羣		
4081	翱	1	1	82FD	7FFA	翱		
4103	莖	1	1	8354	831D	莖		
4111	莖	1	1	835C	8493	莖		
4115	藁	1	1	8360	85C1	藁		
4132		7	7	8371			血𠂇𠂇	眾，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4134	衛	35	35	8373	885E	衛		
4138	表	1	1	8377	88A0	表		
4146	覩	1	1	83A1	89A9	覩		
4184	迹	4	4	83C7	8FF9	迹		
4194	鈐	1	1	83D1	9241	鈐		
4198	鑛	1	1	83D5	945B	鑛		
4220	穎	1	1	83EB	9834	穎		
4235	駢	1	1	83FA	9AA9	駢		
4258	麇	3	3	8452	9E90	麇		
4404	昏	1	1	8547	5B74	昏		
4447	捏	1	1	8572	63D1	捏		
4507	吴	168	168	85D0	5434	吴		
4530	咬	2	2	85E7	6648	咬		
4740	菱	1	1	875D	8593	菱		Unicode 呈現差異。
4778	鞞	1	1	87A5	97B4	鞞		
4902	充	1	1	8862	5156	充		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4904	尅	4	4	8864	5C05	尅		
4908	届	1	1	8868	5C4A	届		
5015	寗	1	1	88F5	5BD7	寗		Unicode 呈現差異。
5046	彝	1	1	8955	5F5C	彝		
5126	暮	2	2	89C7	8B29	暮		
5133	憇	2	2	89CE	6199	憇		
5154	璿	4	4	89E3	7468	璿		
5289	慤	50	50	8ACD	6164	慤		
5463	岢	1	1	8BDE	65F9	岢		
5553	櫟	2	2	8C79	6A52	櫟		
5559	欸	2	2	8CA1	6B35	欸		
5560	槁	2	2	8CA2	6A99	槁		
5591	杞	4	4	8CC1	233CC		木𠂇巳	
5629	淦	1	1	8CE7	6D24	淦		
5706	敍	5	5	8D75	654D	敍		