

《游宦紀聞》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
2008/6/26 陳維君 製作

一、《游宦紀聞》一書(游宦紀聞.xml)使用舊版造字 79 個，字頻 194 次，詳如附件一。這 79 個造字中，68 個可轉成 Windows XP 能顯示的字，字頻 157 次；另外 11 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、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」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 2260 的「癩」字，Unicode 編碼為 3FD7，由於 Windows XP 並不支援，仍須使用構字式「疒△癩」。

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、《游宦紀聞》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	6	6	FBEC			𠄎	
297	𠄏	6	6	FBED			𠄏	
299	𠄑	1	1	FBEF			𠄑	
300	𠄒	1	1	FBF0			𠄒	
752	𠄔	1	1	FEDD			𠄔	
814	𠄖	2	2	8E5C	87C7	𠄖		𧈧，異體字問題。
899	𠄘	1	1	8ED3	99DE	𠄘		
994	𠄚	1	1	8F73	8E75	𠄚		𧈩，異體字問題。
1187	𠄜	1	1	90B9	76D6	𠄜		𧈫，異體字問題。
1205	𠄞	1	1	90CB	770E	𠄞		
1237	𠄠	1	1	90EB	77DF	𠄠		
1278	𠄢	1	1	9155	788D	𠄢		𧈭，異體字問題。
1416	𠄤	1	1	9242	7AE2	𠄤		
1893	𠄦	2	2	9548	886E	𠄦		𧈯，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
2085	穗	1	1	966B	69F5	穗		
2260	癩	1	1	977D	3FD7		疒△頰	
2651	鸚	2	2	99EC	9E1C	鸚		
2748	托	1	1	9AB0	6754	托		
3238		1	1	9DC3			形西广艹糸	西△縻，異體字問題。
3779	霑	1	1	814A	290F1		雨△沱	
3789	冲	3	3	8154	51B2	冲		沖，異體字問題。
3802	劓	1	1	8161	529A	劓		
3843	堦	1	1	81AC	5826	堦		
3852	媼	1	1	81B5	5A63	媼		
3864	峯	5	5	81C1	5CEF	峯		峰，異體字問題。
3883	徧	8	8	81D4	5FA7	徧		
3884	忽	4	4	81D5	6031	忽		
3920	皓	1	1	81F9	6667	皓		
3921	皙	2	2	81FA	6673	皙		
3924	朞	2	2	81FD	671E	朞		期，異體字問題。
3935	槩	8	8	8249	69E9	槩		概，異體字問題。
3966	烟	4	4	8268	70DF	烟		煙，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
3968	煨	2	2	826A	7145	煨		
3974	牀	1	1	8270	7240	牀		
3980	獾	1	1	8276	7328	獾		
3996	踈	4	4	82A8	758E	踈		
4010	臯	1	1	82B6	7690	臯		
4019	矻	1	1	82BF	77F4	矻		
4034	穉	1	1	82CE	7A49	穉		
4038	竈	1	1	82D2	7AC8	竈		
4039	竚	1	1	82D3	7ADA	竚		
4042	笋	1	1	82D6	7B0B	笋		
4044	筋	2	2	82D8	7B6F	筋		
4068	綫	1	1	82F0	7DAB	綫		線，異體字問題。
4080	羣	2	2	82FC	7FA3	羣		群，異體字問題。
4090	脚	4	4	8347	811A	脚		腳，異體字問題。
4107	葱	1	1	8358	8471	葱		蔥，異體字問題。
4118	藜	1	1	8363	8602	藜		
4129	虬	2	2	836E	866C	虬		
4132		13	13	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
4146	覩	1	1	83A1	89A9	覩		
4158	譌	2	2	83AD	8B4C	譌		Unicode 呈現差異。
4180	臯	3	3	83C3	8FA0	臯		
4184	迹	3	3	83C7	8FF9	迹		
4193	鈎	3	3	83D0	920E	鈎		鈎，異體字問題。
4195	鉢	2	2	83D2	9262	鉢		
4207	霧	1	1	83DE	9736	霧		
4240	髯	1	1	8440	9AF4	髯		
4256	鷄	5	5	8450	9DC4	鷄		
4267	鼈	1	1	845B	9F08	鼈		
4322	備	1	1	84B4	4FCC	備		
4450	勑	4	4	8575	524F	勑		Unicode 呈現差異。
4493	特	3	3	85C2	3E40		牛 𠂔 字	特，異體字問題。
4507	吳	25	25	85D0	5434	吳		吳，異體字問題。
4566	礪	1	1	864C	78F5	礪		
4603	憇	1	1	8671	6187	憇		
4679	塏	1	1	86DF	5818	塏		
4778	鞞	1	1	87A5	97B4	鞞		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
4794	藁	1	1	87B5	8448	藁		
4903	𪛗	1	1	8863	5C00	𪛗		
4904	尅	2	2	8864	5C05	尅		
5008	幪	1	1	88EE	5E5E	幪		
5019	畧	3	3	88F9	7567	畧		略，異體字問題。
5302	昏	2	2	8ADA	662C	昏		昏，異體字問題。
5517	榼	3	3	8C55	6985	榼		Unicode 呈現差異。
5542	鑠	4	4	8C6E	93C1	鑠		
5590	朶	1	1	8CC0	6736	朶		朶，異體字問題。
5854	幺	1	1	C6CD	5E7A	幺		
5860	无	3	3	C6D3	65E0	无		