

《澗水燕談錄》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
2008/5/12 葉冠孜 製作

一、《澗水燕談錄》一書(澗水燕談錄.xml)使用舊版造字 75 個，字頻 209 次，詳如附件一。這 75 個造字中，61 個可轉成 Windows XP 能顯示的字，字頻 171 次；另外 14 個字必須轉成構字式，字頻 38 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 3864 的「峯」係「峰」字之異體，故以「峰」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 3846 的「壻」字，Unicode 字型呈現為「壻」，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 2377 的「誓」字，Unicode 編碼為 279AC，由於 Windows XP 並不支援，仍須使用構字式「工几言」。

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、《灑水燕談錄》造字分析表

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
38		1	1	FA65			王厶已	玨，異體字問題
296	厶	2	2	FBEC			厶	
297	厶	4	4	FBED			厶	
675	食	3	3	FE6E			食	
770	扌	1	1	FEEF			扌	
777	宀	2	2	FEF6			宀	
1024	虵	5	5	8FB3	8675	虵		
1067	棊	2	2	8FDE	68CA	棊		
1076	高	1	1	8FE7	5368	高		Unicode 呈現差異
1170	瞳	1	1	90A8	7583	瞳		
1205	眇	3	3	90CB	770E	眇		
1394	啗	1	1	91EB	35D6		口厶啗	啗，異體字問題

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
1646	脰	1	1	93AD	8117	脰		
1893	袞	1	1	9548	886E	袞		
2377	詈	4	4	9855	279AC		形工几言	
2602	鰐	5	5	99BB	9C10	鰐		
2935	穈	1	1	9BCE	7A6A	穈		
3789	冲	1	1	8154	51B2	冲		沖，異體字問題
3797	凭	1	1	815C	51ED	凭		憑，異體字問題
3804	効	1	1	8163	52B9	効		效，異體字問題
3814	却	6	6	816D	5374	却		卻，異體字問題
3824	叙	1	1	8177	53D9	叙		敘，異體字問題
3829	咏	1	1	817C	548F	咏		
3846	壻	5	5	81AF	58FB	壻		Unicode 呈現差異
3858	寃	1	1	81BB	5BC3	寃		冤，異體字問題
3864	峯	3	3	81C1	5CEF	峯		峰，異體字問題
3868	卮	1	1	81C5	5DF5	卮		
3883	徧	3	3	81D4	5FA7	徧		
3888	惠	1	1	81D9	60B3	惠		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
3891	慙	1	1	81DC	6159	慙		慙，異體字問題
3893		1	1	81DE			亻 厶 妄	佞，異體字問題
3897	慙	1	1	81E2	615C	慙		
3932	稜	1	1	8246	6936	稜		
3935	槩	2	2	8249	69E9	槩		概，異體字問題
3974	牀	5	5	8270	7240	牀		
3996	疎	2	2	82A8	758E	疎		
4028	稟	1	1	82C8	7980	稟		稟，異體字問題
4038	竈	1	1	82D2	7AC8	竈		灶，異體字問題
4045	箒	1	1	82D9	7B92	箒		
4080	羣	19	19	82FC	7FA3	羣		群，異體字問題
4090	脚	3	3	8347	811A	脚		腳，異體字問題
4112	蔴	1	1	835D	8534	蔴		
4131	劔	1	1	8370	8842	劔		
4132		14	14	8371			血 厶 亼	眾，異體字問題
4134	衛	17	17	8373	885E	衛		衛，異體字問題
4142	裛	2	2	837B	8943	裛		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4146	覩	3	3	83A1	89A9	覩		
4158	譌	1	1	83AD	8B4C	譌		Unicode 呈現差異
4176	輒	4	4	83BF	8F19	輒		
4184	迹	6	6	83C7	8FF9	迹		
4193	鈎	3	3	83D0	920E	鈎		
4195	鉢	2	2	83D2	9262	鉢		
4198	鑛	1	1	83D5	945B	鑛		
4242	鬪	3	3	8442	9B2D	鬪		Unicode 呈現差異
4262	麪	6	6	8456	9EAA	麪		
4460	勅	1	1	85A1	52D1	勅		敕，異體字問題
4496	段	1	1	85C5	53DA	段		段，異體字問題
4507	吳	10	10	85D0	5434	吳		吳，異體字問題
4735	攏	1	1	8758	58E0	攏		
4768	妬	2	2	8779	59AC	妬		妒，異體字問題
4957	崑	1	1	88BB	5D53	崑		
5074	尫	1	1	8971	5C2B	尫		尫，異體字問題
5131	窻	1	1	89CC	7ABB	窻		窗，異體字問題

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
5176		1	1	89F9			氵厶冗	沉，異體字問題
5225	卹	1	1	8A6B	460F		血厶卩	卹，異體字問題
5290	携	1	1	8ACE	643A	携		攜，異體字問題
5315	浚	1	1	8AE7	3CC2		氵厶公	
5542	鑠	1	1	8C6E	93C1	鑠		
5552	弄	1	1	8C78	5F06	弄		
5558	龕	3	3	8C7E	9E84	龕		
5582	鞣	1	1	8CB8	97B5	鞣		
5590	朶	3	3	8CC0	6736	朶		朶，異體字問題
5591	杞	2	2	8CC1	233CC		木厶巳	杞，異體字問題
5706	敘	9	9	8D75	654D	敘		敘，異體字問題
5752	摠	2	2	8DC5	63D4	摠		