

《湛然居士文集》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
2008/4/21 陳建安 製作

一、《湛然居士文集》一書(湛然居士文集.xml)使用舊版造字 164 個，字頻 565 次，詳如附件一。這 164 個造字中，139 個可轉成 Windows XP 能顯示的字，字頻 453 次；另外 25 個字必須轉成構字式，字頻 112 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 5516 的「○○束」字，原字為「棘」。
- 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「血𠂇𠂇」係「眾」字之異體，故以「眾」字取代。又編號 3797 的「凭」字，係「憑」字之簡體字，亦以標準字的「憑」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 961 的「柎」字，Unicode

編碼為 3B4A，由於 Windows XP 並不支援，仍須使用構字式「木
 𣎵片」。


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、《湛然居士文集》造字分析表

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
79	𣎵	1	1	FAB0	276C2		𣎵 𣎵固	𣎵，校對問題，舊版漢籍錯字，可用程式全部取代。
296	𣎵	12	12	FBEC			𣎵	
297	𣎵	18	18	FBED			𣎵	
299	𣎵	1	1	FBEF			𣎵	
300	𣎵	1	1	FBF0			𣎵	
746	𣎵	1	1	FED7			𣎵	
750	𣎵	3	3	FEDB			𣎵	
766	𣎵	1	1	FEEB			𣎵	
770	𣎵	1	1	FEEF			𣎵	
893	𣎵	1	1	8ECD	5CE9	𣎵		
899	𣎵	1	1	8ED3	99DE	𣎵		
932	𣎵	1	1	8EF4	7683	𣎵		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
938	燮	1	1	8EFA	7215	燮		
961	柎	1	1	8F52	3B4A		木𠂔片	
979	蕙	1	1	8F64	56CF	蕙		
1002	鳩	1	1	8F7B	9D02	鳩		
1205	眎	1	1	90CB	770E	眎		
1240		1	1	90EE			𠂔革卅𠂔伐	鞭，校對問題，舊版漢籍錯字，可用程式全部取代。
1383	糠	1	1	91E0	7A45	糠		
1394	啗	1	1	91EB	35D6		口𠂔啗	啗，異體字問題。
1467	纂	2	2	9275	25C67		𠂔𠂔秦	
1553	媧	1	1	92ED	7179	媧		
1565	繇	2	2	92F9	7DDC	繇		
1784	萁	2	2	9478	855A	萁		
1885	蠱	1	1	9540	882D	蠱		
1893	袞	5	5	9548	886E	袞		
1981	讐	2	2	95C2	8B90	讐		
2044	踈	1	1	9642	8E08	踈		
2046	踪	2	2	9644	8E2A	踪		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
2168	疇	1	1	96E0	91BB	疇		
2502	颺	1	1	98F4	98C8	颺		
2505	翻	2	2	98F7	98DC	翻		
2831	嗞	1	1	9B44	55DE	嗞		
3020	禘	3	3	9C64			禘	
3160	坟	1	1	9D53	575F	坟		
3169	砣	1	1	9D5C	7818	砣		
3770	个	1	1	8141	4E2A	个		個，異體字問題。
3788	慤	1	1	8153	22871		慤  心	慤，異體字問題。
3789	冲	17	17	8154	51B2	冲		沖，異體字問題。
3791	况	4	4	8156	51B5	况		況，異體字問題。
3797	凭	2	2	815C	51ED	凭		憑，異體字問題。
3804	効	2	2	8163	52B9	効		
3805	勅	5	5	8164	52C5	勅		
3809	匱	1	1	8168	5333	匱		
3813	邛	1	1	816C	536D	邛		
3817	廁	1	1	8170	53A0	廁		廁，異體字問題。
3820	厦	2	2	8173	53A6	厦		廈，異體字問題。

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
3824	叙	1	1	8177	53D9	叙		敘，異體字問題。
3829	咏	4	4	817C	548F	咏		
3840	坂	1	1	81A9	5742	坂		
3858	寃	1	1	81BB	5BC3	寃		冤，異體字問題。
3864	峯	18	18	81C1	5CEF	峯		峰，異體字問題。
3876	廻	1	1	81CD	5EFB	廻		迴，異體字問題。
3883	徧	4	4	81D4	5FA7	徧		
3884	忽	4	4	81D5	6031	忽		
3886	悞	3	3	81D7	609E	悞		
3891	慙	3	3	81DC	6159	慙		慚，異體字問題。
3909	擡	1	1	81EE	64E1	擡		
3924	朞	2	2	81FD	671E	朞		
3929	黎	5	5	8243	68C3	黎		
3931	椀	11	11	8245	6900	椀		
3932	椶	1	1	8246	6936	椶		
3935	槩	1	1	8249	69E9	槩		概，異體字問題。
3937	檝	3	3	824B	6A9D	檝		
3942	毡	1	1	8250	6BE1	毡		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
3946	汹	2	2	8254	6C79	汹		
3955	澗	1	1	825D	3D4E		氵 ㄩ 間	澗，異體字問題。
3968	煨	1	1	826A	7145	煨		
3974	牀	10	10	8270	7240	牀		
3976	犁	2	2	8272	7282	犁		
3978	猪	1	1	8274	732A	猪		Unicode 呈現差異。
3989	璿	1	1	82A1	74C8	璿		
3996	疎	18	18	82A8	758E	疎		
4000	疴	1	1	82AC	75B4	疴		
4008	癯	1	1	82B4	766F	癯		
4014	蓋	1	1	82BA	8462	蓋		蓋，異體字問題。
4016	着	13	13	82BC	7740	着		著，異體字問題。
4028	稟	2	2	82C8	7980	稟		稟，異體字問題。
4039	疇	5	5	82D3	7ADA	疇		
4055	粃	1	1	82E3	7C83	粃		
4067	綉	1	1	82EF	7D89	綉		繡，異體字問題。
4077	罇	2	2	82F9	7F47	罇		
4079	待	1	1	82FB	38E5		彳 ㄩ 待	

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4080	羣	26	26	82FC	7FA3	羣		群，異體字問題。
4081	翱	2	2	82FD	7FFA	翱		
4082	耆	1	1	82FE	8008	耆		
4085	耻	2	2	8342	803B	耻		
4089	脇	1	1	8346	8107	脇		
4090	脚	2	2	8347	811A	脚		
4105	菓	7	7	8356	83D3	菓		
4107	葱	3	3	8358	8471	葱		
4114	藥	1	1	835F	8616	藥		
4115	藁	3	3	8360	85C1	藁		
4118	藜	1	1	8363	8602	藜		
4129	虬	5	5	836E	866C	虬		
4132		46	46	8371			血𠂇𠂇	眾，異體字問題。
4134	衛	11	11	8373	885E	衛		衛，異體字問題。
4136	衰	3	3	8375	887A	衰		
4140	衽	1	1	8379	88B5	衽		
4142	裛	1	1	837B	8943	裛		
4146	覩	6	6	83A1	89A9	覩		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4158	譌	1	1	83AD	8B4C	譌		Unicode 呈現差異。
4167	賚	1	1	83B6	8CEB	賚		
4173	駝	2	2	83BC	8EAD	駝		
4176	輒	1	1	83BF	8F19	輒		
4178	輓	2	2	83C1	8F2D	輓		
4184	迹	18	18	83C7	8FF9	迹		
4193	鈎	18	18	83D0	920E	鈎		
4195	鉢	7	7	83D2	9262	鉢		
4218	菲	1	1	83E9	97EE	菲		
4226	凜	4	4	83F1	98E1	凜		
4240	鬚	4	4	8440	9AF4	鬚		
4242	鬪	6	6	8442	9B2D	鬪		Unicode 呈現差異。
4254	鷓	1	1	844E	9D7E	鷓		
4256	鷄	1	1	8450	9DC4	鷄		
4257	鴉	2	2	8451	9D39	鴉		
4260	麋	1	1	8454	9E9E	麋		
4267	鼈	3	3	845B	9F08	鼈		
4307	罵	1	1	84A5	99E1	罵		罵，異體字問題。

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4348	惚	1	1	84CE	506C	惚		
4385	礧	2	2	84F3	255B5		石厶意	
4428	躡	1	1	855F	8E70	躡		躡，異體字問題。
4437	廩	1	1	8568	5EEA	廩		廩，異體字問題。
4450	劊	2	2	8575	524F	劊		劊，異體字問題。
4455	劊	7	7	857A			僉厶刀	
4460	勅	1	1	85A1	52D1	勅		
4496	段	2	2	85C5	53DA	段		
4507	吳	14	14	85D0	5434	吳		吳，異體字問題。
4515	呪	1	1	85D8	546A	呪		
4607	狫	1	1	8675	3E60		犛厶屯	
4702	場	2	2	86F6	5872	場		
4735	壠	7	7	8758	58E0	壠		
4768	妬	1	1	8779	59AC	妬		妒，異體字問題。
4779	鞫	1	1	87A6	97B2	鞫		
4805	瀆	1	1	87C0	51DF	瀆		瀆，異體字問題。
4817	寶	1	1	87CC	5BF3	寶		寶，異體字問題。
4907	橐	1	1	8867	69D6	橐		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4927	岝	1	1	887B	5C9E	岝		
4957	崑	1	1	88BB	5D53	崑		
4981	鳧	1	1	88D3	9CEC	鳧		
4996	斲	3	3	88E2	65B5	斲		
5019	畧	1	1	88F9			畧	略，異體字問題。
5034	弃	1	1	8949	5F03	弃		Unicode 呈現差異。
5043	礦	1	1	8952	5F4D	礦		
5163	憂	3	3	89EC	621E	憂		
5241	鶩	3	3	8A7B	9D5E	鶩		鵝，異體字問題。
5289	慤	2	2	8ACD	6164	慤		慤，異體字問題。
5306		1	1	8ADE			𠄎 𠄎 閒	簡，異體字問題。
5338	暎	1	1	8AFE	668E	暎		
5360	鄺	3	3	8B55	913D	鄺		
5434	湟	4	4	8BC1	6E7C	湟		湟，異體字問題。
5437	并	2	2	8BC4	5E77	并		并，異體字問題。
5478	飴	9	9	8BED	4D77	飴		
5482	璽	3	3	8BF1	8812	璽		
5516		1	1	8C54			∞ 束	棘，校對問題，舊版漢籍錯字，可用程式

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
								全部取代。
5541	蠹	7	7	8C6D	8827	蠹		
5555	迥	2	2	8C7B	9008	迥		
5565	斛	3	3	8CA7	469F		角𠩺𠩺	
5590	朶	1	1	8CC0	6736	朶		朶，異體字問題。
5704	穉	9	9	8D73	6C0A	穉		
5706	敍	11	11	8D75	654D	敍		敍，異體字問題。
5742	澁	1	1	8DBB	6F81	澁		
5768	濶	1	1	8DD5	6FF6	濶		闊，異體字問題。
5797	灑	2	2	8DF2	7054	灑		