

《揭傒斯全集》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
2008/5/21 陳維君 製作

一、《揭傒斯全集》一書(揭傒斯全集.xml)使用舊版造字 183 個，字頻 982 次，詳如附件一。這 183 個造字中，151 個可轉成 Windows XP 能顯示的字，字頻 830 次；另外 32 個字必須轉成構字式，字頻 152 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 4750 的「卅」字，原字應為「卅」。
- 2、標記問題：標記問題主要是漢籍電子文獻在舊轉新的過程中，採用了新的標記語言，而在修改標記時，對於某些標題句做了增刪的動作，因為這些標題句的內容包含了缺字，所以這些更改標記的動作造成 txt 檔與 xml 檔缺字頻次不符的情形。如編號 3804 的「効」字，xml 檔在 256 頁新增了一個標題，因而造成頻次不同，處理方式為將所有頻次皆修改為「効」。
- 3、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 881 的「飈」係「飈」字之異體，故以「飈」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。
- 4、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 1925 的「羈」字，Unicode 字型呈現為「羈」，實際上仍為同一字。

5、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 1001 的「王𠄎斤」。

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 1359 的「襪」字，Unicode 編碼為 25725，由於 Windows XP 並不支援，仍須使用構字式「衤𠄎冀」。

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

四、txt 檔與 xml 檔頻次不同且造字分析表「備註」欄有備註問題情況者，詳細取代內容可參考附件二的手動取代表。附件二的手動取代表的欄位說明如下：

甲、編號：Big5 造字空間為 6217 個，編號由 1 到 6217。

乙、造字：舊版造字字形。

丙、取代：當實際使用字形與造字不同時，在此標出。

丁、檔案原文摘錄：摘錄檔案(揭傒斯全集.xml)中包含該造字的文句段落，其中紅字底線的部分，為需要手動取代的原文。文句後方標示(全部取代)者，表示執行全部取代的動作，不一一列舉所有原文。

戊、手動取代結果：變更過後的檔案內容，其中藍字底線的部分，為修改後的結果。文句後方標示(全部取代)者，表示執行全部取代的動作，不一一列舉所有原文。

附件一、《揭傒斯全集》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	𠄏	20	20	FBED			𠄏	
298	𠄐	3	3	FBEE			𠄐	
299	𠄑	1	1	FBEF			𠄑	
300	𠄒	1	1	FBF0			𠄒	
336	畝	1	1	FC55			畝	
525	膏	2	2	FD75			膏	
763	𠄓	1	1	FEE8			𠄓	
777	𠄔	1	1	FEF6			𠄔	
820	烟	2	2	8E62	70F1	烟		
841	礪	1	1	8E77	7932	礪		礪，異體字問題。
848	髻	3	3	8E7E	9AE4	髻		
881	颺	3	3	8EC1	98C7	颺		颺，異體字問題。
893	峩	6	6	8ECD	5CE9	峩		峨，異體字問題。
899	馳	1	1	8ED3	99DE	馳		
910	嶺	3	3	8EDE	5D78	嶺		嶺，異體字問題。
938	燮	1	1	8EFA	7215	燮		燮，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
949	枅	1	1	8F46	678F	枅		
969	楯	21	21	8F5A	7250	楯		
972	犇	5	5	8F5D	7287	犇		
985	愁	3	3	8F6A	6197	愁		愁，異體字問題。
1000		3	3	8F79			王△巳	玨，異體字問題。
1001		1	1	8F7A			王△斤	待造字。
1010	玨	2	2	8FA5	73CF	玨		玨，異體字問題。
1039	稟	1	1	8FC2	69C0	稟		
1056	陔	1	1	8FD3	9641	陔		
1057	夔	1	1	8FD4	3687		夔	
1076	高	1	1	8FE7	5368	高		Unicode 呈現差異。
1108	糞	1	1	9048	6A86	糞		
1170	瞳	2	2	90A8	7583	瞳		
1359	襪	1	1	91C8	25725		襪△冀	
1371	嶧	1	1	91D4	37FD		山△票	
1382	畱	2	2	91DF	3F66		然△田	
1544	絙	1	1	92E4	7D59	絙		
1565	繇	4	4	92F9	7DDC	繇		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
1612	杼	11	11	9369	6888	杼		
1646	脗	1	1	93AD	8117	脗		
1781	葢	3	3	9475	854B	葢		
1809	困	1	1	94B3	56E6	困		
1916	蚤	1	1	955F			𧈧 𧈨 𧈩 𧈪	
1925	羈	1	1	9568	8989	羈		Unicode 呈現差異。
1926	覓	5	5	9569	8994	覓		覓，異體字問題。
1981	讐	4	4	95C2	8B90	讐		讎，異體字問題。
2148	麇	1	1	96CC	8B4D	麇		
2168	疇	3	3	96E0	91BB	疇		
2492	頽	2	2	98EA	983A	頽		頽，異體字問題。
2502	飈	3	3	98F4	98C8	飈		飆，異體字問題。
2516	餽	1	1	9943	9919	餽		
2564	剌	1	1	9973	34E8		夾 𠂇 𠂈	
2592	贓	2	2	99B1	8D1C	贓		
2663	嶮	6	6	99F8	5DCE	嶮		
2875	亼	1	1	9B70	4EBC	亼		
3780	儻	1	1	814B	5101	儻		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
3784	兪	2	2	814F	516A	兪		兪，異體字問題。
3788	慤	3	3	8153	22871		殼△心	慤，異體字問題。
3797	凭	1	1	815C	51ED	凭		憑，異體字問題。
3802	劓	1	1	8161	529A	劓		
3804	効	4	5	8163	52B9	効		効，標記問題，增加標題，人工取代。
3805	勅	12	12	8164	52C5	勅		
3807	驗	1	1	8166	9A10	驗		
3814	却	46	46	816D	5374	却		卻，異體字問題。
3816	厓	7	7	816F	5393	厓		
3825	疊	5	5	8178	53E0	疊		疊，異體字問題。
3829	咏	2	2	817C	548F	咏		
3830	咤	1	1	817D	54A4	咤		
3834	嗥	2	2	81A3	5637	嗥		
3843	堦	1	1	81AC	5826	堦		
3846	壻	4	4	81AF	58FB	壻		Unicode 呈現差異。
3851	姹	2	2	81B4	59F9	姹		
3852	媼	2	2	81B5	5A63	媼		
3858	寃	4	4	81BB	5BC3	寃		寃，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
3864	峯	77	77	81C1	5CEF	峯		峰，異體字問題。
3868	卮	4	4	81C5	5DF5	卮		
3869	帀	1	1	81C6	5E00	帀		
3876	廻	2	2	81CD	5EFB	廻		迴，異體字問題。
3883	徧	10	10	81D4	5FA7	徧		
3884	忽	2	2	81D5	6031	忽		
3886	悞	1	1	81D7	609E	悞		
3888	惠	2	2	81D9	60B3	惠		
3915	斲	4	4	8459	6585	斲		
3916	斲	2	2	81F5	656D	斲		
3933	桀	1	1	8247	6998	桀		
3940	鬱	1	1	824E	6B1D	鬱		鬱，異體字問題。
3966	烟	29	29	8268	70DF	烟		煙，異體字問題。
3974	牀	9	9	8270	7240	牀		
3975	牕	3	3	8271	7255	牕		Unicode 呈現差異。
3976	犁	1	1	8272	7282	犁		犁，異體字問題。
3977	狗	5	5	8273	72E5	狗		
3978	猪	1	1	8274	732A	猪		豬，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
3979	猫	2	2	8275	732B	猫		貓，異體字問題。
3980	狻	1	1	8276	7328	狻		
3981	珉	2	2	8277	73C9	珉		
3989	璿	1	1	82A1	74C8	璿		
3994	畊	1	1	82A6	754A	畊		
3996	疎	31	31	82A8	758E	疎		
4008	癯	2	2	82B4	766F	癯		
4010	阜	3	3	82B6	7690	阜		
4011	皞	1	1	82B7	76A1	皞		
4012	鼓	6	6	82B8	76B7	鼓		鼓，異體字問題。
4016	着	4	4	82BC	7740	着		著，異體字問題。
4025	礪	1	1	82C5	7934	礪		
4034	穉	1	1	82CE	7A49	穉		
4039	竚	1	1	82D3	7ADA	竚		
4042	笋	1	1	82D6	7B0B	笋		
4044	筋	1	1	82D8	7B6F	筋		
4052	簾	1	1	82E0	7C34	簾		
4055	粃	2	2	82E3	7C83	粃		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
4057	莖	14	14	82E5	585F	莖		莖，異體字問題。
4080	羣	104	104	82FC	7FA3	羣		羣，異體字問題。
4081	翱	2	2	82FD	7FFA	翱		
4090	脚	4	4	8347	811A	脚		脚，異體字問題。
4091	臯	6	6	8348	81EF	臯		
4097	芭	6	6	834E			艹 厶 巳	芭，異體字問題。
4103	苙	1	1	8354	831D	苙		
4107	葱	7	7	8358	8471	葱		葱，異體字問題。
4114	蘘	2	2	835F	8616	蘘		
4115	藁	2	2	8360	85C1	藁		
4129	虬	1	1	836E	866C	虬		
4131	劬	1	1	8370	8842	劬		
4132		79	79	8371			血 厶 爪	眾，異體字問題。
4134	衛	13	13	8373	885E	衛		衛，異體字問題。
4138	表	4	4	8377	88A0	表		
4145	羈	13	13	837E	898A	羈		Unicode 呈現差異。
4146	覩	18	18	83A1	89A9	覩		
4173	駝	2	2	83BC	8EAD	駝		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
4178	輓	1	1	83C1	8F2D	輓		
4184	迹	20	20	83C7	8FF9	迹		
4186	遡	1	1	83C9	9061	遡		
4188	邨	1	1	83CB	90A8	邨		
4193	鈎	1	1	83D0	920E	鈎		鈎，異體字問題。
4203	隣	19	19	83DA	96A3	隣		鄰，異體字問題。
4213	鞞	2	2	83E4	97BE	鞞		
4226	凜	3	3	83F1	98E1	凜		
4242	鬪	8	8	8442	9B2D	鬪		Unicode 呈現差異。
4251	鰕	1	1	844B	9C15	鰕		
4256	鷄	9	9	8450	9DC4	鷄		
4278	亘	4	4	8466	4E98	亘		
4307	罵	2	2	84A5	99E1	罵		罵，異體字問題。
4359	棧	1	1	84D9	6827	棧		
4400	堯	1	1	8543	5160	堯		
4428	躡	2	2	855F	8E70	躡		
4439		1	1	856A			𨇗足已 心	蹠，異體字問題。
4450	翔	1	1	8575	524F	翔		Unicode 呈現差異。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
4460	勅	1	1	85A1	52D1	勅		
4479	嶽	1	1	85B4	3802		山 ㄩ 教	
4566	礪	4	4	864C	78F5	礪		
4750	卅	1	1	8767			卅	**，舊版漢籍錯字， 可用程式全部取代。
4768	妬	5	5	8779	59AC	妬		妒，異體字問題。
4805	瀆	1	1	87C0	51DF	瀆		瀆，異體字問題。
4806	噉	1	1	87C1	56B1	噉		
4844		1	1	87E7			炆 ㄩ 凡	榮，異體字問題。
4907	橐	1	1	8867	69D6	橐		橐，異體字問題。
5013	汜	2	2	88F3			汜	
5019	畧	1	1	88F9			畧	略，異體字問題。
5065	忼	1	1	8968	5FFC	忼		
5103	腮	2	2	89B0	43F0		月 ㄩ 忍	
5176		1	1	89F9			汜 ㄩ 冗	沉，異體字問題。
5241	鶯	1	1	8A7B	9D5E	鶯		鶯，異體字問題。
5378	巉	2	2	8B67	5DD9	巉		
5396	鎔	2	2	8B79	938B	鎔		
5437	并	38	39	8BC4	5E77	并		標記問題，增加標 題，人工取代以標準

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
								字「并」。
5442	栞	1	1	8BC9	67BF	栞		
5447	躡	1	1	8BCE	8E83	躡		躡，異體字問題。
5466	倏	8	8	8BE1	5010	倏		倏，異體字問題。
5488	霽	1	1	8BF7	4A34		雨𠄎對	
5512	𪗗	1	1	8C50	8C3A	𪗗		
5541	蠹	6	6	8C6D	8827	蠹		蠹，異體字問題。
5549	藁	1	1	8C75	85F3	藁		
5555	迥	2	2	8C7B	9008	迥		迥，異體字問題。
5559	欸	2	2	8CA1	6B35	欸		欸，異體字問題。
5565	𪗗	3	3	8CA7	469F		角𠄎𠄎	
5569	瓊	5	5	8CAB	74AE	瓊		
5583	殮	3	3	8CB9	98F1	殮		殮，異體字問題。
5587	歛	5	5	8CBD	6B58	歛		
5588	恠	1	1	8CBE	6061	恠		
5591	杞	3	3	8CC1	233CC		木𠄎巳	杞，異體字問題。
5666	污	8	8	8D4D	6C5A	污		污，異體字問題。
5697	𪗗	7	7	8D6C	7151	𪗗		𪗗，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
5704	𧯛	1	1	8D73	6C0A	𧯛		
5705	廐	2	2	8D74	5ED0	廐		Unicode 呈現差異。
5706	敍	22	22	8D75	654D	敍		敍，異體字問題。
5742	澁	1	1	8DBB	6F81	澁		
5770	汧	1	2	8DD7			汧 𧯛 汧	標記問題，增加標題，人工取代以標準字「汧」。

附件二、《揭傒斯全集》手動取代表

編號	造字	取代	檔案原文摘錄	手動取代結果
3804	効		<u>息</u> (全部取代)	<u>効</u> (全部取代)
5437	并	并	<u>泉</u> (全部取代)	<u>并</u> (全部取代)
5770	汧	汧	<u>完</u> (全部取代)	<u>汧</u> (全部取代)