

《大慧普覺禪師宗門武庫》舊版造字轉碼說明

中研院資訊所文獻處理實驗室
中央研究院語言所文獻語料小組
2010/03/26 丁致伶

轉碼主要工作是把檔案中的舊版造字轉換成 Windows XP 能支援的 Unicode 字形，Unicode 目前共收錄漢字 70194 個字，而 XP 只能支援 20902 個字（詳如表一），不支援之字將以構字式表達。例：造字編號 4121 的「懼」字，Unicode 編碼是 39AC，由於 XP 並不支援，仍需使用構字式「忄 厶羅」。

表一、Unicode 的字數及編碼區段

Unicode	新增字數	新增編碼區段	總字數	WinXP
1.1 版	20902	4E00-9FFF	20902	支援
3.0 版	6582	3400-4DFF	27484	不支援
3.1 版	42710	20000-2A6D6	70194	不支援

一、舊版造字轉碼分析：

《大慧普覺禪師宗門武庫》使用舊版造字 49 個，字頻 188 次，這 49 個造字中，47 個可轉成 Windows XP 能顯示的字，字頻 186 次；另外 2 個字必須轉成構字式，字頻 2 次。

轉碼完成製作轉碼分析表，請參考附件一《大慧普覺禪師宗門武庫》轉碼分析表，欄位說明如下：

- (一) 編號：Big5 造字空間為 6217 個，編號由 1 到 6217。
- (二) 造字：舊版造字。
- (三) 頻次：舊版造字在文件的出現次數。
- (四) Big5：造字的 Big5 碼。
- (五) Unicode：造字所對應的 Unicode 碼。

(六) WinXP：造字在 Windows XP 的對應字形。

(七) 構字式：Windows XP 無法對應字形改用構字式。

(八) 備註凡例：備註欄中記錄轉碼後字形及修改原因，凡例如下：

1. 異體字問題：為了使用者查詢和使用的方便，在處理異體字時最主要的方式是以標準字取代，除非是專有名詞或特殊情形，如：人名、地名等。例：造字編號 5453 的「涅」字，是「涅」的異體字，以標準字「涅」取代。
2. Unicode 字型呈現差異：Unicode 字型與舊版造字有些微差異，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 4450 的「翺」字，Unicode 字型呈現為「翺」，實際上仍為同一字。

附件一、《大慧普覺禪師宗門武庫》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
1023	調	1	8FB2	8AB7	調		
1151	劄	1	9073	5273	劄		
1400	窓	2	91F1	7A93	窓		
1491	擷	2	92AF	6527	擷		
1536	麤	3	92DC	9E81	麤		
1999	贊	6	95D4	8CDB	贊		
2386	瑯	11	985E	7458	瑯		
2409	闕	1	9875	95A7	闕		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
2565	霏	1	9974	9714	霏		
3801	剗	2	8160	5257	剗		
3814	却	52	816D	5374	却		
3817	廁	1	8170	53A0	廁		廁，異體字問題。
3824	叙	2	8177	53D9	叙		
3858	寃	1	81BB	5BC3	寃		寃，異體字問題。
3864	峯	10	81C1	5CEF	峯		
3900	挖	2	81E5	62D5	挖		
3909	擡	4	81EE	64E1	擡		
3944	冰	1	8252	6C37	冰		
3979	猫	2	8275	732B	猫		
3984	瑠	1	827A	7460	瑠		
3990	甄	1	82A2	750E	甄		
3992	嘗	20	82A4	751E	嘗		
3996	疎	1	82A8	758E	疎		
4041	豎	2	82D5	7AEA	豎		
4090	脚	19	8347	811A	脚		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4121	懼	1	8366	39AC		忄 厶 羅	
4146	覩	1	83A1	89A9	覩		
4195	鉢	7	83D2	9262	鉢		
4203	隣	1	83DA	96A3	隣		鄰，異體字問題。
4216	鞞	1	83E7	97C8	鞞		
4220	穎	2	83EB	9834	穎		
4240	鬚	2	8440	9AF4	鬚		
4267	鼈	1	845B	9F08	鼈		
4381	備	1	84EF	511E	備		
4447	捏	1	8572	63D1	捏		
4448		2	8573			形 纟 厂 墨 口	纏，異體字問題。
4450	狃	1	8575	524F	狃		Unicode 字型呈現差異。
4515	呪	1	85D8	546A	呪		
4546	鴟	4	85F7	9D76	鴟		
4768	妬	1	8779	59AC	妬		
5008	僕	2	88EE	5E5E	僕		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5290	携	2	8ACE	643A	携		
5300	鬪	1	8AD8	9B2A	鬪		
5358	龕	1	8B53	23BDB		𠄎牙女 厂毛𠄎	
5453	涅	1	8BD4	23D40		𠄎氵日 工𠄎	涅，異體字問題。
5555	迥	1	8C7B	9008	迥		
5590	朶	1	8CC0	6736	朶		
5702	厨	1	8D71	53A8	厨		
5753	潜	2	8DC6	6F5C	潜		

二、手動取代表說明：

《大慧普覺禪師宗門武庫》檔案中出現「●」字形，共計17字，比對原書後以Unicode字形或構字式手動取代，並製作「●」字手動取代表，請參考附件二。欄位說明如下：

1. 頁碼：「●」字形所在位置之原書頁碼。
2. 檔案原文摘錄：摘錄《大慧普覺禪師宗門武庫》檔案有「●」字形的文句。
3. 原書字形：「●」在原書中之字形，以此字取代檔案原文的「●」。
4. 備註：記錄修改相同詞句的次數或其他事項。

附件二、《大慧普覺禪師宗門武庫》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 943	敢當此記●	荊	
p. 945	在搯●堆上	扌厶垂	
p. 945	僧名法●	門厶豕	
p. 945	僧云法●	門厶豕	
p. 945	住住法●	門厶豕	
p. 945	●於言下	門厶豕	
p. 945	●一日持	門厶豕	
p. 945	●不諾	門厶豕	
p. 945	遂問●曰	門厶豕	
p. 945	爾試說看·●	門厶豕	
p. 946	也好●	灋	
p. 946	●·圓悟	灋	
p. 950	不是·僧●	扌厶麼	
p. 950	狗子●	灋	
p. 952	曰兜率●	灋	
p. 953	●若爾不	口厶力	
p. 955	有關●	檢	