

《郁迦羅越問菩薩行經》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
2009/04/28 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 4696 的「燿」字，Unicode 編碼是 3E0C，由於 XP 並不支援，仍需使用構字式「火𠂇霍」。

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：

《郁迦羅越問菩薩行經》使用舊版造字 17 個，字頻 40 次，這 17 個造字中，16 個可轉成 Windows XP 能顯示的字，字頻 39 次；另外 1 個字必須轉成構字式，字頻 1 次。

轉碼完成製作轉碼分析表，請參考附件一《郁迦羅越問菩薩行經》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 4768 的「妬」字，是「妒」的異體字，以標準字「妒」取代。

附件一、《郁迦羅越問菩薩行經》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3805	勅	2	8164	52C5	勅		
3814	却	4	816D	5374	却		
3858	寃	1	81BB	5BC3	寃		寃，異體字問題。
3992	嘗	1	82A4	751E	嘗		
4028	稟	1	82C8	7980	稟		
4134	衛	13	8373	885E	衛		衛，異體字問題。
4195	鉢	2	83D2	9262	鉢		
4260	麀	1	8454	9E9E	麀		
4696	燿	1	86F0	3E0C		火𠄎霍	
4738	睽	2	875B	777A	睽		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4768	妬	1	8779	59AC	妬		妒，異體字問題。
4775	妣	4	87A2	59B7	妣		
5133	憇	1	89CE	6199	憇		
5300	鬪	3	8AD8	9B2A	鬪		
5316	晉	1	8AE8	664B	晉		晉，異體字問題。
5453	涅	1	8BD4	23D40		形: 日 工	涅，異體字問題。
5472	瓦	1	8BE7		瓦		

二、手動取代表說明：

《郁迦羅越問菩薩行經》檔案中出現「●」字形，共計6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《郁迦羅越問菩薩行經》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、《郁迦羅越問菩薩行經》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 25	見乞●者	𠂇	有 4 處
p. 26	沓●說法	𠂇	
p. 29	神捷沓●	𠂇	