《尹文子》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
2008/3/5 陳建安 製作

一、 《尹文子》一書(尹文子.xml)使用舊版造字 8 個，字頻 23 次，詳如附件一。這 8 個造字中，6 個可轉成 Windows XP 能顯示的字，字頻 21 次；另外 2 個字必須轉成構字式，字頻 2 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、 備註凡例：

1、標記問題：由於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，對於某些內容做了增刪的動作，因為這些內容包含了缺字，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。針對標記問題，只需要將 xml 檔中剩下的舊漢籍造字，一一修改為對應的 Unicode，若無對應的 Unicode，則以構字式取代。如編號 4134 的「衛」，因在書末的校勘記中另外引用了 1 次原文，故在全書中出現 2 次，但由於 xml 檔的標記方式已將校勘記插入原文中，不再另外於書末標出，因而在 xml 檔中只出現 1 次。處理的方法是將 xml 檔中剩下的舊漢籍造字，一一修改為對應的 Unicode，又以「衞」字為「衛」之異體，故以「衛」字取代。

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」係「眾」字之異體，故以「眾」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 4242 的「鬬」字，

Unicode 字型呈現為「鬬」，實際上仍為同一字。

三、 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《尹文子》手動取代表。附件二《尹文子》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、 取代：當實際使用文字與造字字形不同時，在此標出。
丙、 檔案原文摘錄：摘錄檔案(尹文子.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。
丁、 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。

四、 《尹文子》中未確定字形「●」共 1 處，詳細處理內容可參考附件三的未確定字形「●」手動取代表。附件三《尹文子》未確定字形「●」手動取代表欄位說明如下：

甲、 檔案原文摘錄：摘錄檔案(尹文子.xml)中包含「●」的文句段落，其中紅字底線的部分，為需要手動取代的原文。
乙、 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。

附件一、《尹文子》造字分析表

編號	造字	頻次(txt)	頻次(xml)	Big5	Unicode	WinXP	構字式	備註
3935	槩	1	1	8249	69E9	槩		概，異體字問題。
3996	踈	4	4	82A8	758E	踈		
4080	羣	13	13	82FC	7FA3	羣		群，異體字問題。
4132		1	1	8371			血厽从	眾，異體字問題。
4134	衛	2	1	8373	885E	衛		標記問題,異體字問題，改為衛。
4184	迹	1	1	83C7	8FF9	迹		

編號	造字	頻次(txt)	頻次(xml)	Big5	Unicode	WinXP	構字式	備註
4242	鬬	1	1	8442	9B2D	鬬		Unicode 呈現差異。
5306		1	1	8ADE			竹厶閒	簡，異體字問題。

附件二、《尹文子》手動取代表

編號	取代	檔案原文摘錄	手動取代結果
4134	衛	鬒有鰥夫	衛有鰥夫

附件三、《尹文子》未確定字形「●」手動取代表

檔案原文摘錄	手動取代結果
而耳不可以●	而耳不可以目从祭