《太平惠民和劑局方》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
2008/4/9 黃榮順 製作

一、 《太平惠民和劑局方》一書(太平惠民和劑局方.xml)使用舊版造字 27 個，字頻 169 次，詳如附件一。這 27 個造字中，23 個可轉成 Windows XP 可顯示的字，字頻 123 次；另外 4 個字必須轉成構字式，字頻 46 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、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 3817 的「廁」係「廁」字之異體，故以「廁」字取代。又如編號 3966 的「烟」係「煙」字之簡體，故以「煙」字取代。

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。所以如造字編號 4979 的「錫」字，Unicode 編碼為 88D1，由於 Windows XP 並不支援，仍須改用構字式「食 ⿰易」。

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、 「太平惠民和劑局方.xml」缺字外發現之錯字共 1 處，修正情形詳列附件二，附件二「太平惠民和劑局方.xml」缺字外錯字修改表體例說明如下：

甲、 檔案原文摘錄：摘錄檔案(太平惠民和劑局方.xml)中包含該錯字的文句段落，其中紅字底線的部分，為需要手動取代的原文。

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。若為原書所無，則删去之。

附件一、「太平惠民和劑局方.xml」造字分析表

編號	造字	字頻(txt)	字頻(xml)	Big5	Unicode	WinXP	構字式	備註
660	丐	42	42	FE5F			丐	
685	礻	1	1	FE78			礻	
687	犭	1	1	FE7A			犭	
792	帟	1	1	8E46	5E2C	帟		
1067	棊	1	1	8FDE	68CA	棊		棋，異體字問題。
1387	暨	19	19	91E4	7796	暨		
2021	趣	1	1	95EA	8D82	趣		
2325	瞷	9	9	97E0	77A4	瞷		
3255	善	2	2	9DD4	5584	善		

編號	造字	字頻(txt)	字頻(xml)	Big5	Unicode	WinXP	構字式	備註
3817	廁	3	3	8170	53A0	廁		廁，異體字問題。
3883	徧	1	1	81D4	5FA7	徧		
3966	烟	8	8	8268	70DF	烟		煙，異體字問題。
4000	疴	1	1	82AC	75B4	疴		
4042	笋	2	2	82D6	7B0B	笋		
4063	蘖	18	18	82EB	7CF5	蘖		
4195	鉢	8	8	83D2	9262	鉢		
4263	麴	38	38	8457	9EAF	麴		
4270	鼐	1	1	845E	9F45	鼐		
4501	猜	1	1	85CA	7356	猜		
4525	唄	2	2	85E2	54EF	唄		
4616	壳	2	2	867E	58F3	壳		
4902	兖	1	1	8862	5156	兖		
4964	妓	1	1	88C2	59AD	妓		
4979	餳	2	2	88D1	296FF		食 ⿰ 易	
5074	厖	1	1	8971	5C2B	厖		

編號	造字	字頻(txt)	字頻(xml)	Big5	Unicode	WinXP	構字式	備註
5477	腴	1	1	8BEC	815D	腴		
5704	氊	1	1	8D73	6C0A	氊		

附件二、「太平惠民和劑局方.xml」缺字外錯字取代表

檔案原文摘錄	手動取代結果
?犭⿰⿱⿰⿱胃皮	犭⿰⿱⿰⿱胃皮